A TREATISE

ON

ELECTRICAL THEORY

AND THE

PROBLEM OF THE UNIVERSE.

A TREATISE

ON

ELECTRICAL THEORY

AND THE

PROBLEM OF THE UNIVERSE.

CONSIDERED FROM THE PHYSICAL POINT OF VIEW,
WITH MATHEMATICAL APPENDICES.

BY

G. W. DE TUNZELMANN, B.Sc.,

MEMBER OF THE INSTITUTION OF ELECTRICAL ENGINEERS;
FORMERLY PROFESSOR OF PHYSICS AND ASTRONOMY, H.M.S. "BRITANNIA,"
DARTMOUTH.

Illustrated.

LONDON:
CHARLES GRIFFIN AND COMPANY, LIMITED,
PHILADELPHIA: J. B. LIPPINCOTT COMPANY.
1910.

Dedicated

IN GRATEFUL RECOGNITION TO THE AUTHOR'S FORMER TEACHERS AND PRESENT FRIENDS, PRINCIPAL G. CAREY FOSTER, F.R.S., AND PROFESSOR O. HENRICI, F.R.S., AND TO THE MEMORY OF ANOTHER FORMER TEACHER AND FRIEND OF HIS STUDENT YEARS, THE LATE PROFESSOR W. K. CLIFFORD, F.R.S.

PREFACE.

The primary aim of this treatise is to develop as complete a presentation of the existing state of electrical theory as can be effected without demanding, on the part of the reader, any further mathematical attainments than the elementary acquaintance with simple methods which may be regarded as the minimum equipment of the student of physics.

The subject has therefore been developed from the physical, rather than the mathematical, point of view, and, except where it has been possible to arrive at the required results by quite simple and elementary mathematical methods, physical illustrations and explanations have been employed in place of analytical demonstrations. In the few cases where this has not been possible, the results obtained by means of mathematical analysis have been stated in as simple and untechnical a form of words as was possible.

I believe that the course adopted will make the work of use to a wider range of physical students than if it had been addressed only to readers familiar with modern mathematical analysis, and that it will give the electrical engineer sufficiently clear physical concepts of the actions in the electromagnetic field to enable him to apply them to the design of electrical apparatus and machinery, and to avoid the errors, and consequent loss of time, which have frequently resulted from the lack of them. Although there are no mathematical difficulties in the text, it has been found impossible, without unduly limiting the scope of the treatise, to avoid the inclusion of occasional sections containing somewhat difficult reasoning in reference to considerations of an abstract character. These sections have been printed in large type, so that they may be deferred to a second reading, or omitted altogether by those who find them too difficult.

Mathematical as well as unmathematical readers should derive benefit from the physical method of presentation adopted, and, in order to increase the value of the work to the former class, the more important mathematical points arising in connection with the text have been dealt with in appendices. Some of these

simply supply proofs, not otherwise available in an easily accessible form, of statements made in the text. Others contain supplementary information which will be of value to the reader, or considerations designed to facilitate his study of more advanced treatises and of original papers. Others, again, deal with difficulties and controversial questions of theory. With the exception of Appendix K, and portions of H and M, which necessarily presuppose somewhat high mathematical attainments on the part of the reader, they should present no serious difficulty to any student of mathematical physics.

The modern development of electrical theory has extended its purview over the whole field of physical phenomena, with the possible exception of gravitation and cohesion; and although these phenomena may not be definitely included within the electromagnetic scheme, there is shown to be good reason for believing them to be closely related to it. A complete survey of the field of electric theory, therefore, necessarily involves the consideration of "The Problem of the Universe," so that no apology is required for the title of the book.

From this point of view the subject is of very wide and general interest, and I have therefore kept in view the requirements of the general reader, with only a rudimentary knowledge of physics, by special attention to simplicity in the descriptive portions, with somewhat ampler explanations than would sometimes have been requisite for the serious physical student alone. The first two chapters will be found of great assistance by readers of this class, enabling them to refer at once to definitions of physical quantities and descriptions of phenomena which may not be familiar to them.

Throughout the work the consideration of experimental details has been avoided. These details are easily accessible elsewhere, and their introduction would have tended to obscure the development of the argument, in addition to the very practical objection that it would have increased the size, and therefore the cost, of the book to an inordinate extent. The two introductory chapters are therefore not intended to provide the general reader, entirely ignorant of the subject, with a sufficient knowledge of the experimental basis of electrical theory. Such readers should consult some elementary introduction to the subject, such as Professor Andrew Jamieson's *Practical Elementary Manual of Electricity and Magnetism.* They, and other readers also, will find these chapters useful for reference, and for the exact definitions, with some at least of which they may not be familiar.

The general reader is advised to confine his attention in the first place to the chapters and appendices marked * in the Table of

Contents, and afterwards to turn his attention to such portions of the remaining chapters as he finds interest him without demanding more serious study than he is inclined to devote to the subject. He may, moreover, if he be interested only in the descriptive matter, also omit Chapters III. and V., which deal, in simple, untechnical language, with general fundamental considerations of great intrinsic interest and importance.

The serious student of electrical theory cannot do better than study Foster and Porter's *Elementary Treatise on Electricity and Magnetism* as an introduction to, and in conjunction with, the present work. The reader interested in the historical development and fundamental principles of dynamics will find Mach's *Principles of Mechanics* of the utmost value. He may also study with advantage *La Science et l'Hypothèse* and *La Valeur de la Science*, by H. Poincaré, both of which are obtainable in English. The mathematical reader will not fail to study Sir Joseph Larmor's *Æther and Matter*, and those of sufficient mathematical attainments will find themselves amply repaid by the study of his great series of memoirs on "A Dynamical Theory of the Electric and Luminiferous Medium," which extend to some three hundred pages of the *Philosophical Transactions* for 1894–5–7. My own indebtedness to these papers is immense. They may be regarded, in fact, as constituting the framework which has determined the form of a large part of this volume. I may warn the reader of an error, to which Sir Joseph Larmor was kind enough to call my attention during the progress of this work, in the investigation of the pressure of radiation in these memoirs. A modified and correct investigation will be found in *Æther and Matter*. Chapter VII. will, for the student, naturally form an introduction to Sir J. J. Thomson's standard work on *The Conduction of Electricity through Gases*, which contains full details of the experimental work as well as of the mathematical development.

In Chapter VIII. I have attempted the very difficult task of presenting, in a form which will be intelligible to the ordinary physical student possessing only an elementary mathematical equipment, a connected outline of the Faraday-Maxwell theory as developed in Maxwell's great treatise. I trust this chapter, in conjunction with the corresponding mathematical appendices, may be found of use to the mathematical reader as an introduction and guide to the study of Maxwell. Jeans' Treatise on *The Mathematical Theory of Electricity and Magnetism* will be found a valuable help in its fuller exposition of the mathematical analysis employed by Maxwell. Jeans' work was written for the fairly advanced mathematical student, while Maxwell's work was written for mathematicians. The chapter on the "Electron

Theory" is the hardest chapter in the book, and, although much of it is perfectly simple, some portions will necessitate hard, continuous thinking on the part of the reader to whom the subject is new. A great deal of labour was spent upon it in order to simplify the presentation as far as possible without shirking the inherent difficulties of the subject. The reader should not be discouraged if he finds it necessary to read the more difficult portions of this chapter a second or even a third time before fully grasping their significance. A careful study, during the first reading, of the diagrammatic illustrations (pp. 201–2) of the twist in the Kelvin model of the ether corresponding to the action in the electromagnetic field during current induction will be found to make the second reading of the chapter very much easier. A preliminary reading of Poynting and Thomson's *Heat*, referred to in a footnote on p. 233, will be found a great aid to the comprehension of the general energetic considerations dealt with in the latter part of Chapter X. and in subsequent chapters.

Mathematical readers are advised to study Drude's *Optics*, which has been translated into English, in connection with Chapter XII. Other works which will be of value to non-mathematical as well as to mathematical readers are Professor Schuster's *Theory of Optics* and R. W. Wood's *Physical Optics*.

Chapter XIV., on "Metallic Conduction and Radiation," in the form first sent to the publishers, was the least satisfactory and one of the shortest chapters in the book. Sir J. J. Thomson had apparently disproved the earlier form of the theory, and provisionally replaced it by one which appeared to me to be too artificial in its assumptions to be regarded as more than a temporary hypothesis which might prove of value in leading the way to a more satisfactory one. As the exposition of both these theories is to be found in Sir J. J. Thomson's *Corpuscular Theory of Matter*, it seemed unnecessary to do more than briefly refer to them. The publication in the *Philosophical Magazine* for June 1909, before the chapter had been set up in type, of Professor J. H. Jeans' important paper on "The Motion of Electrons in Solids," which is largely drawn upon in Appendix M, enabled me to withdraw it, and to rewrite it on the basis of the generalisation of the old theory, which was there shown to be completely satisfactory. Mathematical readers will naturally refer to Jeans' *Dynamical Theory of Gases*, which is so frequently referred to in this chapter. Readers who wish for more information on the experimental details than is to be found in the chapter on the "Kinetic Theory of Matter" in Poynting and Thomson's *Heat* may refer with advantage to Meyer's *Kinetic Theory of Gases*, which contains

a very complete and readable experimental presentation of the subject, with some mathematical appendices. The proofs of the law of equipartition given in the latter are the old and defective ones.

For further information on radioactivity the student will naturally refer to Professor Rutherford's standard treatise on the subject, which contains full experimental details and such little mathematical investigation as is required. Soddy's work on *Radioactivity* and the Hon. R. J. Strutt's *Becquerel Rays* will also be of value. The general reader will find a very simple presentation of the subject in Makower's volume in the International Science Series.

The work of writing Chapter XXIII. was greatly facilitated by the numerous references to original papers in Professor J. Zenneck's article "Gravitation" in the *Encyklopädie der Mathematischen Wissenschaften*, vol. VI. I have endeavoured throughout to give the name of the publication as well as the date of all such original papers as, from lateness of date or otherwise, are not to be found in the standard works referred to, but when they are to be found there I have usually only mentioned the date, in order to avoid the undue multiplication of footnotes.

That the absence of aberration is the essential point in the consideration of the speed of propagation of gravitational attraction was originally suggested to me by Professor C. V. Boys, F.R.S., in conversation, and on subsequent reference to Laplace's well-known investigation considered on page 420, there remained no doubt in my mind that this is the correct view.

I have attempted, on page 443, to illustrate the legitimacy of employing the perfectly well-defined mathematical concept of four-dimensional space which is involved in the hypothesis that gives the simplest representation of gravitational attraction as an action transmitted through the ether. A further illustration, and one which may open up an unexplored line of thought to the mathematical reader, has been suggested to me by a remark made by Sir Oliver Lodge, that time might perhaps be regarded as a sort of fourth dimension to space. Some of the observations which follow will probably be somewhat obscure to the non-mathematical reader, but he will find no difficulty in grasping their general import.

Our thinking process is almost entirely confined within the limits of time and space. The highly elaborated symbolic logic of the mathematician enables him only very partially to emancipate himself from these conditions, and the non-mathematician, generally, if not always, finds himself entirely incapable of transcending these limits. To the evolutionist the

explanation appears fairly obvious. These modes of thinking provide us with the simplest possible representation of the external world, and one which is amply sufficient for the practical purposes of existence. They have therefore formed for untold ages a portion of the nature of the human race, gradually ascending from their less developed form in its animal ancestry. They have been felt as limitations only within the few thousand years in which men have attempted to extend their conceptions of the external world beyond the necessities of animal existence into the wider domain of philosophical thought. Even at the present time, and in the most civilised communities, those who attempt to think for themselves still form an almost infinitesimal proportion of the whole. A much larger proportion, it is true, now provide themselves with something which they regard as a mental outfit, not infrequently pieced together out of concepts as mutually incongruous as the materials of a jackdaw's nest. There is therefore little need for wonder that we should require all the aid obtainable from mathematical analysis to effect even the most partial emancipation from fetters which were once instruments of progress. Sir William Rowan Hamilton's conception of algebra as the science of pure time was, like everything emanating from the mind of that great mathematical genius, extremely suggestive, but Riemann's concept of space as a manifold was even more so. It was this conception which definitely legitimised the concept of generalised space which has been so fruitful both in the development of pure mathematics and in its applications to the more recondite physical problems. I remember, in my student days, discussing Hamilton's concept of algebra with one of the most illuminative of mathematical teachers, Professor W. K. Clifford, in the light of Riemann's work, and suggesting to him that I could see no reason against a similar extension of the uni-dimensional manifold of time, and his observation that in that case we ought to be able to send an angel into the future who should return and inform us of what it had in store for us. The extension of Hamilton's concept to the higher complex algebras of comparatively recent development would necessarily involve this extension of time also into a multi-dimensional manifold. Moreover, Riemann's conception of space as a manifold may be directly extended to time, which, as the concept is employed by us, is simply a uni-dimensional manifold.

Sir Oliver Lodge's observation has suggested to me a way of regarding the question which will, I think, be apprehensible even to the non-mathematician. Let us imagine a point to be moving with a finite speed, which, for simplicity, may be supposed constant, for a finite time, and without change of direction. Its path will then

be a straight line, and the position of the point on this line at any given instant may be considered as the *present* of the moving point, while the parts of the line on opposite sides of the point may be regarded as representing its past and its future history respectively. Each of these parts will form a uni-dimensional manifold representing a flux of time, and, from the point of view of the *present* position of the point, the first part will be considered as now existing in memory only, and the second part as existing as yet only in imagination. We may look at the same motion in quite a different aspect, by regarding it as represented by a uni-dimensional space-manifold instead of a uni-dimensional time-manifold. Any selected position of the point will then divide this manifold into two parts, both regarded as existing actually, and not merely in memory and imagination, respectively.

Now let the straight line move in a fixed direction perpendicular to itself, but otherwise in the same manner as the point, for a time and at a speed, or through a distance, the same as in the former one. Every intermediate position of the line will divide a uni-dimensional manifold, having straight lines as its elements, into two parts. If the extension is considered as represented in time, these two parts, representing the past and the future history of the selected line, will again be considered as existing, the one in memory, the other in imagination. If, on the other hand, the extension is represented in space, both parts will be regarded as having present existence. In each case the complete trace of the line will form a square. Now let the square be supposed to move, in a direction perpendicular to its plane, for the same time at the same speed, or through the same distance, as before. The complete trace will form a cube consisting of a uni-dimensional manifold of planes, and any intermediate plane may be regarded in either the time or the space aspect.

If we now consider the whole process of arriving at the cube, we find that, considered as an extension in space, it may be regarded as a continuous one. The motion of the point traces out a line, the motion of the resulting line traces out a square, and the motion of the square traces out a plane. The square forms a two-dimensional manifold of points, while the cube forms a three-dimensional manifold of points, all the points being capable of being regarded as existing simultaneously, since the whole motion has been effected without requiring any point to move into a position already occupied. Considered, on the other hand, as an extension in time, we are unable to give the same interpretation to the motion throughout. We find that we have simply twice changed the representation of a time-flux, picturing

it first by the motion of a point, then by the motion of a line, and finally by the motion of a plane.

If we attempt to continue the process by the addition of another stage, we find no difficulty from the point of view of extension in time. We simply have to make another change in the representation of a time-flux, regarding it now as represented by the addition of cube to cube, and it is evident that this may be continued indefinitely. At each stage, however, we only add a fresh unidimensional manifold, incapable of combination with any of the preceding ones. From the point of view of spatial extension we are unable to proceed further, because we are incapable of conceiving points in space into which the points of the existing cube may move without assuming positions already occupied. We could continue the process mathematically without any difficulty, and through as many stages as we pleased, but the human mind is incapable of mentally visualising any of the further stages.

The important points for our consideration are that—(1) throughout every stage which we are capable of visualising, the same system of relations may be represented alternatively as a division between two simultaneously existing systems, or a division between a past which has ceased to exist and a future which has not yet come into existence; (2) mathematical analysis shows beyond the possibility of question that the obstacle to the indefinite continuation of the former process is due entirely to the limitation of our power of visualisation of the relations, and not to anything inherent in the relations themselves.

It would therefore be perfectly legitimate to represent what we call the present state of the universe, not as a division between a past which has ceased to exist and a future which does not yet exist, but as a division between two continuously existing systems. From such a point of view, time would be entirely eliminated. The three-dimensional space in which our minds picture the external world would then form the division between two four-dimensional spaces; and its series of changes, as presented to our consciousness, would be regarded as arising from the motion of the division, our three-dimensional space, through the four-dimensional space in which it is contained. To an intelligence capable of such a visualisation of the containing space, the past and present of our universe would appear as existing simultaneously in the present.

The extension, in Chapter XXIV., of the ordinary methods of physical inquiry to an investigation of the nature of the Primal Intelligence is the logical outcome of the development of electrical theory into an apparently all-embracing theory of the material universe, taken in connection with the fundamentally important

proposition referred to and illustrated in the text and demonstrated in Appendix K. It is there shown to be of infinite mathematical improbability, or, in ordinary language, impossible, that a universe containing molecular matter should constitute a self-maintaining conservative system. Its existence as a conservative system would necessarily involve its continuous guidance by intelligently directed power. The existence of Primal Intelligence as a necessary antecedent to the existence of the universe and to any consistent scheme of scientific knowledge has received ample philosophical demonstration, and I am not aware of its even having been regarded as an open question by any great thinker. Most modern physicists have, however, regarded it as being entirely beyond the reach of physical investigation. This was my own opinion until, in the course of developing the present work, I was led to the solution now presented to my readers in this chapter. I should strongly advise those who, without being philosophers, are interested in the philosophical questions referred to above, to read A. J. Balfour's *Foundations of Belief*, where they are dealt with in language charmingly free from philosophical technicalities, but, at the same time, with convincing logic. I should have referred to the book, and probably quoted from it in the text, except that, through not realising its character and scope, I only read it after the completion of the final proofs. The reader who studies this work from the point of view of Chapter XXIV. will not be troubled with the difficulties which Mr Balfour's somewhat hesitating adoption of the idealistic standpoint might otherwise occasionally give rise to. Another work which should interest readers of this chapter, and which I have referred to in Appendix Q, although the text had been completed before I saw it, is Sir Oliver Lodge's *Life and Matter*. One of the suggestions which I have made on page 469, as possibly throwing light on the question of the freedom or determinism of the will, is of similar character to one which he had made in this little volume. I may also refer them to an interesting paper by Professor W. R. Sorley, "The Interpretation of Evolution."[1] There is also an article on "Directivity" as characteristic of living beings, by Professor Henslow,[2] to which my attention has been called only within the last few days, and which is in close accordance with a portion of the argument of the same chapter. The whole problem of the *directivity* of matter by mind is shown in this chapter to be reducible to the problem of the *directivity* of energy by mind, and it is shown further, that even so complex a case as the guidance of the material universe by the

[1] *British Academy Proceedings*, vol. iv., 1909.
[2] *Hibbert Journal*, vol. vi. p. 157, 1907.

Primal Intelligence is capable of representation in the simple form of the guidance of a moving point along a prescribed path.

The purposive action upon its immediate environment, of the mind which manifests itself in association with the material structure of every living form, may be represented in a similar manner. The greatest of human minds, however, can effect but comparatively small influence upon the moving point characteristic of the system, that is to say, the portion of the universe to which its power extends. The intelligence of such a mind is, moreover, of such a limited nature that even in those simple cases where it is possible to ensure, with reasonable certainty and accuracy, the attainment of a desired end, there is always the possibility that unintended and undesired results may also ensue.

Many readers may be expected, when reading this chapter, to make the comment that, although the directive power of mind over energy, and therefore over matter, has been clearly shown to exist, no explanation has been offered of the nature of the interaction. They will say with du Bois-Reymond, *ignoramus*, we do not know; they may even go further with him and say, *ignorabimus*, we shall never know. Prophecy is, however, notoriously unsafe, and especially negative prophecy. In one, and that the narrowest, sense I have no expectation of his prophecy ever being falsified. In a wider sense I think I can discern indications of its rapidly approaching falsification. It is pointed out in Chapter III. that the representations by which physicists have attempted to systematise and correlate—and explanation can have no other meaning—natural phenomena, have been formed on the model of mechanics, the simplest and earliest developed branch of physics. The explanation of physical phenomena hence came to mean their representation in terms of mechanical concepts, and many able biologists, having no definite conception of what a mechanical representation means, have spoken and written loosely and inaccurately of mechanical representation of the whole of natural phenomena, including even the phenomena of life and mind. The definite meaning of a mechanical representation is given on page 45, and the progress of physical investigation has shown that such concepts can only approximately represent even the simplest of material phenomena. We find that the phenomena of molecular physics are totally incapable of complete representation in mechanical terms. Quasi-mechanical concepts have proved of the utmost value in linking up our knowledge of molecular physics, the second approximation to a representation of the universe, with the mechanical model forming the first approximation. The reader will learn, however, that such concepts are only of value as stepping-stones, and that

if we attempt to carry them forward in their entirety into the molecular scheme we soon find them obstructing instead of assisting our progress.

If, then, we have to abandon the mechanical concepts of the first model as we make our way into the profounder depths of molecular physics, it becomes obvious that we cannot expect any assistance from them in carrying us forward into the mental scheme which begins to dawn upon our vision only when we have reached the innermost recesses of the molecular scheme. Not only, therefore, does the progress of science offer no promise of a mechanical representation of the interaction between mind and matter; it indicates in no uncertain terms the utter futility of any such expectation. The inadequacy of quasi-mechanical concepts of molecular physics, in comparison with the power of the mathematical methods which are gradually unlocking its secret chambers, has led some mathematical physicists to advocate the entire abandonment of the concept of the ether in favour of representations of the observed relations by means of differential equations taking no account of the ether. The arguments in favour of such a course have been presented with clearness and ability in two papers by N. Campbell in the *Philosophical Magazine* for February 1910. Although some of his arguments are, in my opinion, open to criticism, it cannot, I think, be seriously disputed that a simpler scheme of electromagnetic equations would be possible in the absence of the concept of the ether. On the other hand, it hardly appears to me possible that any physicist can suppose that, in the absence of such a concept, we should have arrived at the stage of knowledge of relations between the phenomena of molecular physics which makes it possible to formulate such equations. I believe, moreover, that the reader of the present work will find in its pages ample justification for my conviction that there is not the least indication that the concept is likely to be less useful in further progress towards the unification of our representations of the whole system of nature than it has proved to be in the past.

The conviction that it would be the depth of unwisdom to cast aside, in order to simplify our systems of differential equations, the quasi-mechanical concepts which have proved of so great assistance in attaining our present standpoint, does not in any way prevent my contemplating, not only the possibility, but the probability of the replacement of the existing representation of the universe by one of an even more all-embracing character. To attempt to simplify this representation by abandoning the concept of the ether, and with the ether the variability of mass, would, in my opinion, be a step backward. The concept of the ether has

led us, as I have shown in Chapter XXIV., to the conclusion that energy is a more fundamental concept than either ether or matter. It is therefore more fundamental than the concept of mass, so that the indicated path of progress is not the remodelling of our representation in order to make it capable of simpler expression in terms of a system of dynamics in which mass was regarded as fundamental. What we have to contemplate is, in my opinion, the remodelling of our system of dynamics on the basis of energy in the place of mass. We may then begin to contemplate the ultimate possibility of a future remodelling, in which mind will replace energy as the fundamental basis of the physical scheme.

As our mechanical and quasi-mechanical concepts are already affording evidence of their insufficiency, we have to look forward to the formation of entirely new physical concepts, of the nature of which we can have at present not even the glimmering of an idea. In the meantime we shall have to rely, in the future as in the past, on the guide that has never yet failed us, but has ever proved capable of developing new powers as they have been called for in order to cope with the ever increasing difficulty of the way. Mathematical analysis, the most wonderful of all the products of the human mind, ever continues its growth at an ever increasing rate.

Without venturing to make rash prophecies which time would probably show to be extremely foolish, there is, to my mind, a very deep significance in some of the recent advances in our knowledge of pure mathematics. Just as Galileo, Huygens, Newton, and some of their greater contemporaries and successors effected the complete disentanglement of physics and theology, to the manifest advantage of both, so Riemann, Lobachewsky, Cantor, Peano, and other brillant mathematicians have disentangled pure mathematics from the physics to which it has proved such a powerful aid. The freedom so acquired is already accelerating the growth of the former, and so making it a still more efficient aid to the progress of the latter. The whole system of pure mathematics, as I have pointed out in Chapter V., is now definitely proved to be, as foreshadowed in Boole's wonderful *Laws of Thought*, nothing else than an elaborately developed system of symbolic logic. It is a pure product of the mind, independent of all empirical knowledge, and must therefore, almost of necessity, be fitted for the investigation of mental phenomena.

The intimate relationship found to subsist between purely mental processes and those by which we form our concepts of the external world would lead us to expect that the former, like the latter, would require transformation into simpler approximative concepts, or models, before they would be amenable to the thought

process embodied in mathematical analysis. Considerable progress in the construction of such models has already been made by modern psychologists, the work of psychology consisting in the analysis of the complex mental processes, in the light of experimental research, into distinct portions or elements. These portions are no more distinct in actual experience than are the portions into which the phenomena of the material universe have been similarly analysed. It is, I think, clearly brought out in the present work that elements do not exist in our actual experience of the external world. But it is made equally clear that no progress can be made in the correlation of mental processes until a transformed representation in terms of elements has been effected. When the psychologists have accomplished this preliminary work to a sufficient extent there appears no reason to doubt the power of mathematical analysis to make the initial steps towards a first approximation to a definite mental scheme, or that its development in the requisite directions will fail to maintain its place sufficiently in advance of the progress of experimental psychology to make it always the trustworthy guide which it has proved at lower altitudes.

There now remains only the pleasant duty of expressing my thanks to those who have assisted me in one way or another in the execution of a laborious undertaking, but one which has been of absorbing interest to myself, and which I trust may also interest my readers and prove helpful as a connected outline of the scheme of nature as viewed from the standpoint of a physicist whose aim has been to arrive at the truth and to present it to his readers as adequately as lay in his power.

My obligations to my friend Professor G. M. Minchin, F.R.S., are beyond my powers of adequate expression. He has been kind enough to read nearly the whole of the proofs from the beginning of Chapter III., and besides the correction of misprints which had escaped my notice, and slips and inaccuracies of expression for which I myself was to blame, he has in very many instances pointed out portions of the argument where further explanation was advisable in order to make it more readily intelligible or to add to its precision. His important simplification, and addition to the precision, of magnetic theory, arising out of his reading the proofs, are dealt with in detail in Chapter VIII. and Appendix F.

My old friend and former teacher Principal Carey Foster, F.R.S., to whose suggestion and encouragement the amplification of the work greatly beyond the first restricted plan was largely due, has kindly read the first two chapters in proof, and suggested several simplifications and improvements. Another friend I have

to thank is Sir William Ramsay, F.R.S., both for reading the proofs of Chapters XV. to XVII. on radioactivity, and for information, embodied in the text, with respect to the recent determination in his laboratory of the period of decay of radium. I have also to thank Professor Svanté Arrhenius for his kindness in reading the proofs of Chapters XVIII. to XX., and for several suggestions.

In Chapter XXIV., where, besides attempting to obtain at least a first approximation towards the solution of a problem which has hitherto usually been regarded as beyond the reach of physical methods, I have attempted further to extend the partial solution obtained by the aid of general philosophical considerations, a subject in which I have no claim to be considered an expert, I sought further criticism from philosophical friends, and Professor W. R. Sorley, D.Litt., Professor of Moral Philosophy in the University of Cambridge, Canon Gildea, and Mr A. H. Douglas-Hamilton were all kind enough to read the proofs, and I am indebted to each one of them for valuable suggestions.

I have to thank Dr G. A. Schott for kindly sending me unpublished details of his calculations, and for reading the proofs of Appendices H and P, in which they are dealt with; and I have to thank Sir Joseph Larmor, Secretary of the Royal Society, for information on several points during the progress of the work.

I have been very greatly indebted to Mr White, the Assistant Librarian of the Royal Society, and his assistants, for invaluable assistance, sometimes involving much time and trouble, in tracing some of the numerous memoirs which had to be consulted.

I have to express my thanks to the following authors and publishers for permission to reproduce extracts from the works mentioned:—To the Syndics of the Clarendon Press, in the case of Clerk-Maxwell's treatise on *Electricity and Magnetism*; to Sir Joseph Larmor and the Syndics of the Cambridge University Press, in the case of *Æther and Matter*; to Mr M. E. Darwin and Messrs John Murray, in the case of Charles Darwin's *Climbing Plants* and *Insectivorous Plants*; to Mrs Max Müller and Messrs. Longmans, Green & Co., in the case of Professor Max Müller's works; to Messrs Macmillan & Co., in the case of Professor Huxley's works; to Professor W. R. Sorley and Messrs William Blackwood & Sons, in the case of *The Ethics of Naturalism*; and to Mr Charles Watts, in the case of Professor Haeckel's *Riddle of the Universe*. In conclusion, I have to express my great indebtedness to my own publishers for their unfailing courtesy and unremitting care and attention during the arduous work involved in the correction of the proofs.

G. W. DE TUNZELMANN.

April 18, 1910.

CONTENTS.

The *'s indicate the chapters and appendices most suited to the requirements of the general reader.

*CHAPTER I.

Fundamental Electrical Phenomena.

*CHAPTER II.

Units and Measurement.

*CHAPTER III.

Meaning and Possibility of a Mechanical Theory of Electricity.

*CHAPTER IV.

The Ether.

*CHAPTER V.

The Ether as a Framework for Absolute Motions.

*CHAPTER VI.

Relations Between Ether and Moving Matter.

*CHAPTER VII.

Conduction in Gases and Dielectrics.

CHAPTER VIII.

The Faraday-Maxwell Theory.

CHAPTER IX.

The Electron Theory.

CHAPTER X.

Magnetism and the Dissipation of Energy.

CHAPTER XI.

Contact Electrification and Electrolysis.

CHAPTER XII.

Optical Phenomena.

CHAPTER XIII.

The Mechanism of Radiation.

CHAPTER XIV.

Metallic Conduction and Radiation.

*CHAPTER XIX.

The Solar Corona, the Aurora, and Comets' Tails.

*CHAPTER XX.

Radioactivity in Stars and Nebulæ.

CHAPTER XXI.

Arrangement and Number of Electrons in an Atom.

MATHEMATICAL AND OTHER APPENDICES.

APPENDIX A.

PRINCIPLES OF THE VARIATIONAL METHOD EMPLOYED FOR DERIVING THE EQUATIONS OF EQUILIBRIUM, OR OF MOTION, OF A DYNAMICAL SYSTEM.

APPENDIX B.

HAMILTON'S PRINCIPLE.

APPENDIX C.

MAXWELL'S ANALYTICAL METHODS.

APPENDIX D.

MAXWELL'S ETHER STRESS SYSTEMS.

APPENDIX E.

APPENDIX F.

DETERMINATION OF POTENTIAL GRADIENTS IN THE INTERIOR OF MAGNETS AND OTHER BODIES.

APPENDIX G.

RELATION BETWEEN THE DIELECTRIC CONSTANT AND THE DENSITY OF A FLUID MEDIUM.

APPENDIX H.

THE ALTERNATIVE MECHANICAL REPRESENTATIONS OF ELECTRIC DISPLACEMENT AND MAGNETIC INDUCTION, AND THE MOBILITY OF ELECTRONS.

APPENDIX I.

THERMODYNAMICAL NOTES AND DEFINITIONS.

APPENDIX J.

APPENDIX K.

The Eternal Existence of an Uncontrolled Universe is Infinitely Improbable.

APPENDIX L.

Electromotive Force of Diffusion.

APPENDIX M.

Notes on Professor J. H. Jeans' Contributions to the Electron Theory of Metallic Conduction and Radiation.

APPENDIX N.

Astronomical Anomalies and the Generalisation of Newton's Law of Gravitation.

APPENDIX O.

The Clustering of Gravitational Matter in the Universe.

APPENDIX P.

Hydrodynamical Ether Theories of Gravitation and the Determination of the Scale of Atomic Structure.

* APPENDIX Q.

Haeckel's "Riddle of the Universe."

* APPENDIX R.

A Simple Experiment in Thought Transference, and an Hypothesis by which it may be accounted for.

A TREATISE ON ELECTRICAL THEORY AND THE PROBLEM OF THE UNIVERSE.

CHAPTER I.

FUNDAMENTAL ELECTRICAL PHENOMENA.

If a dry glass rod be rubbed with one of resin or ebonite, and the two rods suspended by means of dry threads of undyed silk in such a manner that they can move freely, they will be found to attract each other; but two rods of ebonite rubbed with glass, or of glass rubbed with ebonite, will repel each other. The rods are then said to be electrified, and the fact that they exhibit two distinct phenomena, viz. attraction and repulsion, shows that there are two distinct kinds of electrification. That of the glass is called vitreous electrification, and that of the ebonite is called resinous electrification. Friction is only one among many methods of producing electrification.

If either of the rubbed rods be brought near to a light body, such as a pith-ball suspended by a dry silk thread, the latter will first be attracted and then repelled. When the ball is repelled by a vitreously electrified rod, it will be attracted by a resinously electrified rod; but after coming in contact with the latter, it will immediately be repelled, and will then be attracted by a vitreously electrified rod. These experiments show that the ball, on coming into contact with the electrified rod, receives an electrification similar to that of the rod, so that electrification is a property which can be transferred from one body to another by contact. That which is transferred is called electricity. It does not follow from the fact that there are two kinds of electrification that there are necessarily two kinds of electricity. Later on we shall see that there are reasons for believing two kinds of electricity to exist, but that in these experiments it is only the resinous electricity that is transferred.

Since a suspended pith-ball acts as an indicator of electrification, it may be called an electroscope; but a much more sensitive electroscope may be made by attaching two similar strips of gold-leaf to the lower extremity of a metal rod supported in a vertical position by means of a dry glass support. As glass has a tendency to condense moisture from the air upon its surface, it should be enclosed in a glass case containing sulphuric acid, which will keep the air dry by absorbing the moisture. The upper end of the metal rod of the gold-leaf electroscope may conveniently be made to terminate in a flat metal plate or disc forming a small table.

If an electrified pith-ball be lifted by its silk thread and laid upon the disc of the electroscope, the gold leaves will be seen to diverge slightly, and this divergence may be increased by laying other similarly electrified balls on the disc, or by allowing a single ball to touch, first an electrified rod, and then the disc of the electroscope, several times in succession. If the ball, after touching the charged rod, be touched by the finger, or by a metal rod held in the hand, before placing it on the disc of the electroscope, it will not increase the divergence of the leaves. This shows that it has lost its electrification; and we see that while the finger and the metal rod allow the electricity to escape, the dry glass support and the dry silk thread do not. The former are called conductors of electricity, and the latter insulators. This is only a rough classification, as the best insulators will allow a charge of electricity to escape in time. Metals are the best conductors, and will discharge an electroscope instantly. The finger will do the same, although more sensitive methods of measurement show it to be a much worse conductor than any metal, while metals, although all good conductors, differ considerably among themselves in this respect. A support of dry glass, free from lead, will retain a charge for many days. Water is a fairly good conductor, which is the reason that the glass and silk thread have to be kept dry. The insulating power of silk is very much less than that of flint glass.

Let a flannel cap be made to fit tightly over the end of an ebonite rod, and attach to it a dry silk thread. Place the cap on the end of the rod, and turn it round several times so as to rub it against the rod. Then bring the end of the rod with the cap on it over the disc of the electroscope. There will be no divergence of the leaves. Remove the cap from the end of the rod by means of the silk thread, and hold, first the end of the rod, and then the cap, over the disc. A divergence of the leaves will take place in each case, but on removing them without allowing them to touch the disc, the leaves will close again completely. This is called electrification by induction. If a second ebonite rod rubbed with glass be brought near the disc when the leaves

are diverged by the first rod, the divergence will be increased; but if this is done when the leaves are diverged by the flannel cap, the divergence will be diminished. The first rod was therefore similarly electrified to the second one, viz. resinously, while the flannel was vitreously electrified. Since either the first rod or the flannel was capable of producing divergence separately, but no divergence was produced by the rod and cap together, it follows that their electrifications were such that their sum was zero. In other words, their relation was that of positive and negative quantities of equal magnitude, so that the friction of the cap on the rod caused no actual generation of electricity, but only a transference of something which was not merely a physical quality, but a physical quantity, whatever its nature might be.

This transference may consist of a flow of positive electricity from the ebonite to the flannel, or of negative electricity from the flannel to the ebonite, or of both these in combination. The experiments, so far, show no reason for selecting as an explanation any one of these three possibilities; but from the point of view of simplicity the first two are preferable to the third, and of these I shall select the second, and assume that what is transferred is negative or resinous electricity, as this is the conclusion which we shall find indicated at a later stage.

If the early experimenters who decided to call vitreous electricity positive had possessed the knowledge which we now have, they would undoubtedly have made resinous electricity positive and vitreous electricity negative. We shall see later on that there is a very large body of evidence pointing to the conclusion that certain small negatively electrified particles which have been shown to act as carriers of the electric current in the phenomena of vacuum-tube discharge, and to be emitted from many of those substances which are known as radioactive, are nothing else than charges of negative electricity, quite free from ordinary matter. These are known as *negative electrons*. The results of the foregoing experiments may therefore be expressed, in terms of this hypothesis, as follows:—

1. When electrification is produced by friction between two bodies a passage of negative electrons takes place from one to the other, so that one loses and the other receives a certain number of negative electrons.

2. A pair of bodies, one of which has gained, and the other lost, negative electrons, will attract one another.[1]

[1] All material substances exert a gravitational attraction upon one another, but this is so small that it may be neglected in comparison with the electrical actions—a course which is more convenient than introducing continually recurring qualifying clauses referring to the gravitational attraction.

3. A pair of bodies, both of which have gained, or lost, negative electrons, repel each other.

4. Some bodies allow the negative electrons to traverse their substance very readily, others only with difficulty. The former class are called good electric conductors, and the latter bad electric conductors, or insulators. If any body could be found which would allow electrons, positive or negative, to traverse its substance without offering any resistance to their passage, it would constitute a perfect conductor; while if any body could be found which would permit of no passage of electrons through its substance, it would constitute a perfect non-conductor, or insulator.

It will be seen that, so far, this hypothesis closely corresponds with Benjamin Franklin's single-fluid theory of electricity. Although it perfectly represents the facts, there is one respect in which it does not appear quite satisfying to the mind, in that it does not offer any explanation of the fact that no electric reactions are exhibited by bodies containing the normal number of electrons except when they are brought into the neighbourhood of bodies which contain an excess or defect, in which case, if the former are conductors, electrical induction takes place. It was to remedy this defect that Symmer proposed the two-fluid theory, which assumed that unelectrified bodies contained equal quantities of positive and negative electricity. Now the existence of positive electrons cannot be considered at present as definitely demonstrated by their experimental isolation, although the recent researches of some physicists have led them to believe that they have succeeded in effecting this. We shall, however, see that the explanation of electrical phenomena on Faraday's principles, which exclude direct action at a distance, and allow only of action taking place directly between contiguous portions of a continuous medium, necessarily involves the association with every negative electron of a corresponding numerically equal positive electron. The experimental facts, however, appear to show that the positive electrons are seldom, if ever, dissociated from matter, or capable of moving freely through solid bodies, the latter property appearing to be confined to a certain proportion of the negative electrons, hence known as *free electrons.* The present point of view is therefore practically a modification of the old two-fluid theory, in which the positive fluid is supposed to be fixed by the material atoms, and only a portion of the negative fluid is mobile.

We may now explain the attractions and repulsions exhibited by electrified bodies by actions between the electrons only, by assuming that electrons of similar kind repel each other, while electrons of opposite kind attract each other. A body in the

normal, or unelectrified, condition will be one which contains as many negative as positive electrons so distributed as to neutralise each other's action. If some of the negative electrons in such a body are displaced by the action of external electrification, without becoming separated from the body, they may no longer neutralise each other's action, and the body will be electrified by induction. For example, when the ebonite rod which had been rubbed by the flannel cap was brought separately near to the disc of the electroscope, the divergence of the leaves would be explained by negative electrons being repelled from the disc to the leaves by the excess of negative electrons on the rod. The flannel cap, on the other hand, would have a deficiency of negative electrons, and therefore an excess of positive ones. It would therefore draw negative electrons from the leaves into the disc, leaving an excess of positive electrons in the leaves, which would consequently diverge again. If, while the influencing body remained near the disc, the latter were connected to the earth or other large conductor, by touching it, for example, with the finger, the free negative electrons would be driven beyond the influence of the electrified rod in the first case, and the leaves would close, owing to the removal of the free negative electrons which had caused their divergence. The disc, being nearer than the leaves to the influencing rod, would contain a smaller proportion of mobile negative electrons; but on the removal of the rod a redistribution would take place, some of those in the leaves passing into the disc, which would therefore share its deficiency in negative electrons with the leaves, so that both leaves and disc would now have a deficiency of negative electrons, and the former would again diverge. The whole electroscope would have acquired a permanent positive charge. If the flannel cap were employed as the influencing body, the leaves would also close when the disc was connected to the earth or other large conductor, because negative electrons would be drawn from the latter until the leaves had become neutral and an excess of negative electrons was held in the disc by the attraction of the flannel. On removing the latter, there would be a redistribution of negative electrons over leaves and disc, by a flow from the latter to the former. The leaves would therefore diverge again, the electroscope remaining with a permanent negative charge.

If two bodies are electrified by friction with each other so as to become equally and oppositely electrified, and are then introduced, without touching it, into a closed metallic vessel resting on the disc of the electroscope, whatever may be their positions within the vessel, there will be no divergence of the leaves. This shows that the vessel has received no electrification. If either one of the

bodies be removed, the leaves will diverge, and they will do so to the same extent whichever body be removed. When the second body is also removed, the leaves will of course close completely. The experiment is due to Faraday.

If the two equally and oppositely charged bodies are conductors (which must of course be supported by insulating threads or rods), and one of them be allowed to touch the inside of the vessel while the other be within but not touching it, no immediate effect will be produced on the leaves; but when both bodies are removed the leaves will diverge, and the body which was allowed to touch the inside of the vessel will be found to have completely lost its charge, while the other will retain its full charge.

If a number of charged conductors, some of which may be positive and some negative, be successively introduced into the closed vessel, allowed to touch it, and then withdrawn, each will be found to have given up the whole of its charge, and the charge of the vessel will therefore be equal to the algebraical sum of the charges on the separate conductors. This gives us a practical method of adding together the electrical charges of a number of bodies. It also shows that the electrification of a conductor in a condition of equilibrium must be entirely confined to the surface.

A simple modification of the last experiment will enable us to obtain a charge equal to any multiple of a fixed charge which may be taken as a provisional unit. Introduce a body bearing the fixed charge into an insulated metallic vessel. Connect the latter to earth for a moment. All signs of electrification will disappear. Withdraw the charged body without allowing it to touch the vessel. The latter will then be charged equally and oppositely to the fixed charge. Then introduce the vessel thus charged into a larger closed insulated vessel and allow it to touch, when it will give up the whole of its charge to the outer vessel. The smaller vessel may then be given a second charge of equal amount, and this second charge given to the larger vessel; and this process may be repeated as often as desired.

If two bodies whose dimensions are small compared with the distance r between them, receive charges e_1 and e_2, the force acting between them will, for a given *dielectric*, that is to say, insulating, medium, be proportional to the product of the charges divided by the square of the distance separating them, viz. to the expression $\frac{e_1 e_2}{r^2}$. This result was first arrived at by Coulomb, by experiments with the torsion balance; but such measurements are not susceptible of much accuracy, and a far more exact proof of the

law is obtained from the consideration that it is a mathematical consequence of the fact that a conductor placed within a hollow closed conductor, and allowed to touch it, gives up the whole of its charge to the latter.

Any area or space within the influence of electrified bodies is called an *electric field*. An electrified body placed anywhere in an electric field will experience a force in virtue of its electrification. If the body and its charge are too small to cause any sensible alteration in the field, this force, acting on the body in any given position, will be proportional to its charge, and its direction will be reversed if the sign of the charge be reversed. If these conditions be not fulfilled, the force will, in general, depend on the charge, and on the size and shape, of the body.

If e be the small charge on the small body, the force acting on it at any point of the field will therefore be of the form $\mathbf{E}e$, where $\mathbf{E}$ is a quantity depending on the electrified bodies to which the field is due. $\mathbf{E}$ is called the *electric force* at the given point of the field, and is printed in heavy type to call attention to the fact that it is a quantity which has direction as well as magnitude, or what is called a *vector quantity*. See also page 115. Since it not only tends to move an electrified body, but to move the electricity within the body (so that the negative electrons would be displaced in the direction opposite to that of $\mathbf{E}$), Maxwell also calls it the *electromotive force at the point*. The total work done on the small body by the electric forces, when it is moved from a given point in the field to a point beyond the influence of the field, will be independent of the path followed if the field is an electrostatic one, viz. one due only to the presence of stationary electrified bodies, and will be proportional to e, so that it may be represented by Ve. The quantity V is then called the *potential* at the given point. If the charge e could be made equal to unity without sensibly disturbing the field, the potential at any point would therefore be measured by the work done on a body charged with a unit of electricity in moving it from that point to an infinite distance. The electric force in any given direction at any point of an electrostatic field is equal to the rate of decrease of the potential at that point due to a displacement in the given direction. This is a simple mathematical consequence of the definition of potential.[1] The direction of the resultant

[1] *Mathematical Note.*—Let r be the distance of a point P from an electrified point charged with a quantity e of electricity, then the potential V at the point P, due to the charge e, will by definition be the work done against the electrical force due to e in bringing a unit charge from an infinite distance up to the point P. This is found by dividing the whole path of the unit charge into elementary portions, expressing the work done in traversing each element as the product of the length of the element into the electric force at the place

force will therefore be that direction in which the rate of change of the potential is greatest. It follows from this that a positively electrified body will tend to move from places of higher positive potential to places of lower positive, or of negative, potential; and a negatively electrified body—a negative electron, for instance—will tend to move in the opposite direction. It follows that when electricity is in equilibrium on a conductor, all parts of the conductor must be at the same potential.

A surface described in an electric field, such that the electric potential is the same at every point of it, is called an *equipotential surface*, or *level surface.* It is evident that the surface of a conductor in equilibrium must be a surface of this kind.

A line traced out by a point moving always in the direction of the force in an electric (or other) field of force is called a line of force. Since the lines of force can have no components along the equipotential surfaces, they must everywhere cut these at right angles; and, as a special case, the lines of force must always enter or leave the surface of a conductor in equilibrium at right angles.

If two conductors at different potentials are connected by a thin conducting wire, the tendency of electricity to flow along the wire will be proportional to the difference in the potentials of the two conductors. This difference of potential is therefore said to be the *electromotive force* between them. Electromotive force

---

occupied by the element, derived from the law of force given above, and then adding together the infinite number of infinitely small quantities of work so obtained. The sum so obtained as the value of V is $\frac{e}{r}$, viz.

$$V = \int_0^\infty \mathrm{E}dr = \int_0^\infty \frac{e}{r^2}dr = \frac{e}{r}\cdot$$

Again, since

$$\frac{dV}{dr} = \frac{d\frac{e}{r}}{dr} = \frac{e}{r^2},$$

it follows that the force $\frac{e}{r^2}$ acting at any point distant r from the electric charge e is equal to the rate of change of V in the direction of r, and the force in any other direction making an angle θ with r is $\frac{dV}{dr}\cos\theta$, which is the rate of change of V in the new direction, which proves the statement in this simple case. For the general proof of the proposition see Appendix E.

For a number of charges e_1, e_2, e_3, etc., at distances r_1, r_2, r_3, etc., from P, we should have

$$V = \frac{e_1}{r_1} + \frac{e_2}{r_2} + \frac{e_3}{r_3} + \text{etc.} = \sum\frac{e}{r}.$$

is a more general term than difference of potential, the latter being only one of the causes which may give rise to electromotive force. The electromotive force E between two points must be carefully distinguished from the electric force **E** at a point, the former being obtained by a summation, along a line joining the two points, of the work done by the latter on a unit charge. In mathematical language, E is the *line-integral* of **E**, taken between the two points as limits.

If one conductor be insulated while all neighbouring conductors are kept at zero potential by being earth connected, the ratio borne by the electric charge on the insulated conductor to its potential is called its *capacity*. A system consisting of a pair of conductors separated by a thin stratum of an insulating material is usually called a condenser. Its capacity is found to be directly proportional to the area of the opposed surfaces, and inversely proportional to the thickness of the insulator. It also depends on the nature of the latter. The ratio of the capacity of a condenser with a given insulator, or dielectric, to the capacity of a similar condenser with air as the dielectric, was called by Faraday the *Specific Inductive Capacity* of the dielectric, while Maxwell called it the *Dielectric Constant.*

It has been known from very ancient times that the iron ore commonly called lodestone possesses some very remarkable properties.

1. If a portion of it be freely suspended at any part of the earth's surface except the magnetic poles, it will, in general, tend to assume a certain definite direction.

2. It has the power of attracting iron very strongly, and in a lesser degree certain other substances, such as the metals nickel and cobalt.

3. These two properties can be communicated to pieces of iron or steel by mere contact, and more effectively by rubbing them in a certain manner with a piece of the lodestone. They also acquire the power of transmitting these properties to other pieces of iron or steel. Bodies possessing these properties are called magnets. Pieces of soft iron magnetised in the above manner, or by other more convenient methods, soon lose their magnetic properties, but pieces of steel are found to retain them.

The ends of a long thin magnet are usually called its *poles*, and the straight line joining them is called the *axis* of the magnet. In the case of an indefinitely long and thin magnet, uniformly magnetised throughout its length, the ends would act as centres of force, and the rest of it would exhibit no magnetic properties. Long, thin, magnetised steel wires approximate to these conditions, and by suspending them from a torsion balance, Coulomb

succeeded in showing that the law of force between two magnetic poles is similar to that between two electric charges, substituting the strength of a pole for the amount of an electric charge. A unit magnetic pole may then be defined as a north-pointing pole which, when placed at unit distance from an equal and similar pole, repels it with the unit of force. A south-pointing pole is taken to be negative. The product of the length of a long, uniformly magnetised, thin magnet into the strength of its positive pole is called its *magnetic moment*.

The neighbourhood of a magnet, or a system of magnets, will constitute a field of magnetic force. The lines of force in any plane section of such a field may be rendered evident in a very simple manner, employed by Gilbert some two hundred years earlier, but the significance of the result was first recognised by Faraday, provided the system can be so arranged as to bring the desired plane into an approximately horizontal position. A sheet of stiff paper is fixed in this horizontal plane, and iron filings are dusted lightly over it through a sieve formed of fine wire-gauze or muslin. Each filing will then become a magnet, under the action of the magnetic force, and will tend to arrange itself with its greatest length in the direction of the line of force at the point where it falls upon the paper. A field of magnetic force which is uniform and parallel throughout the space occupied by a magnet placed in it will only exert a directive action on the magnet, and will not tend to make it move bodily along the lines of force, which shows that the two poles of a magnet must be numerically equal in strength. This is easily verified by floating a magnetised needle on a piece of cork on water, as the earth's magnetic field is sensibly parallel and uniform within the space occupied by the needle. The cork and needle will assume a certain direction, viz. that in which the length of the needle lies as near as possible to the direction of the magnetic force, but there is no motion of the body, as a whole, in any direction.

If a long thin magnet which exhibits no magnetic properties except at the ends be broken into a number of pieces, each piece forms a magnet with equal and opposite poles. If these pieces are all replaced in their original relative positions, they will simply reconstitute the original magnet, showing no magnetic properties except at the two free ends.

This suggested to Weber a theory of magnetisation, or *magnetic induction*, which, as subsequently modified by Hughes, has obtained general acceptation. The modified theory assumes that every molecule of a magnetic substance, *i.e.* of a substance capable of being acted upon by magnetic force, is itself a magnet, but that, in the unmagnetised condition of the substance, these molecular

magnets group themselves into closed circuits or chains. In that case the magnetic actions of the molecules would neutralise one another, so that the body would not act as a magnet. If, however, either pole of a magnet were brought near to it, this pole would attract the dissimilar poles of the molecules, and repel the similar poles, with the result of breaking up the molecular chains, and leaving the molecules free to turn into positions with their magnetic axes more nearly in the direction of the magnetic force. The magnetic actions of the molecules would then reinforce one another, so that the body would be converted into a magnet with the opposite polarity in the portion nearest the pole of the inducing magnet. The intensity of magnetisation of the body, viz. the ratio of its magnetic moment to its volume, would therefore be increased by increasing the magnetising force, and also by any procedure which would increase the mobility of the molecules; for example, by setting them in vibration by mechanical shocks while under the influence of the magnetic force, or by heating the substance, and allowing it to cool under the influence of the magnetic force. These results are in full accordance with the observed facts. The theory receives further confirmation from the changes of shape found to accompany magnetisation, and the slight sounds heard when iron is suddenly magnetised or demagnetised by means of an electric current.

The first definite connection between magnetic and other electric phenomena was established by the Swedish physicist Oersted in the year 1820. Several observers had previously noticed cases in which magnetism appeared to be produced or destroyed in steel needles by electric discharges through them or near them. These led to various conjectures as to the relation between electricity and magnetism, and Oersted at first made a number of unsuccessful attempts to find some relation between magnets and bodies charged with electricity. It was in trying the effect on a magnet of a wire heated by an electric current that Oersted made his great discovery. He soon found that the observed effect was due to the direct action of the current itself, and not to its heating effect. He found that a magnetic needle placed near a wire carrying an electric current tended to set itself at right angles to the wire, with the same end always pointing forward as the magnet was moved round the wire. He described this very appropriately by the phrase "the electric conflict acts in a revolving manner." He also observed that the direction of the needle was reversed when the direction of the current was reversed.

The subject was shortly afterwards taken up by Ampère, who pointed out that the direction of deflection was such that to a person lying along the conductor, facing the magnet, with the

current flowing from his feet to his head, the north pole of the needle would always turn to his left hand. Ampère also discovered that conductors carrying electric currents attract or repel each other just like magnets, and proved experimentally that the magnetic action of a small plane circuit, at distances which are great compared with the dimensions of the circuit, is proportional to that of a magnet whose axis is perpendicular to the plane of the circuit, and whose magnetic moment is equal to the area of the circuit multiplied by the strength of the current. In other words, he showed that the action of the circuit is equal to that of the magnet, provided the unit for measuring the strength of the current is suitably chosen.

We see, then, that the space in the neighbourhood of a conductor carrying an electric current may be considered as a magnetic field. A magnetic field produced by the action of electric currents in conductors is usually called an electromagnetic field, and the magnetic force is sometimes called the electromagnetic force.

The discovery of the mutual action of electric currents and magnets provided the means of detecting the presence of electric currents and comparing them quantitatively. Ampère first employed a compass needle suspended within a vertical loop of wire forming part of an electric circuit as a current detector, and called it a galvanometer. The loop was placed so that the needle lay in its plane when no current was passing. Schweigger improved upon this by suspending the needle within a coil of wire, instead of within a single turn; and in 1837 Pouillet modified the instrument so as to make it capable of exactly comparing the strengths of two currents by means of the respective deflections of the needle. In 1846 Weber devised an instrument in which an electric current was made to traverse in succession a fixed coil and a small movable one suspended within it. This instrument he called an electrodynamometer. It enables a current to be measured by means of the deflection of the suspended coil.

The similarity of the magnetic effects of magnets and electric currents suggested to Ampère that the molecular magnetism of Weber's theory might be accounted for by the assumption that each of the molecules has an electric current circulating round it. The currents would then be brought into approximate parallelism when a body was magnetised. The currents flowing in opposite directions round the adjacent portions of two neighbouring molecules would destroy each other's effects, so that a magnetised rod might be considered as equivalent to a series of currents flowing in the same direction round its surface.

It would therefore resemble a helix [1] carrying an electric current, which is generally known as a *solenoid*. At the time that this theory was propounded it was open to the objection that every conductor, such as we are acquainted with, becomes heated by the passage of an electric current, whereas these molecular currents would have to be supposed devoid of heating effect, as otherwise a magnetic substance would afford a continuous supply of heat, which would be in conflict with experience. This difficulty disappears in the modern view, considered in a later chapter, that these molecular currents are due to the unresisted orbital motions of electrons within the atoms.

It was not until the year 1831, eleven years after Oersted's discovery of the magnetic action of an electric current, that Faraday, who had for some time been trying to produce electric currents by magnetic action, succeeded in discovering the conditions of electro-magnetic induction. Faraday had wound three coils of copper wire, each about twenty-four feet long, and insulated from one another by means of twine and calico, upon a ring of soft iron six inches in external diameter. The free ends of one of the coils were connected with a copper wire passing over a compass-needle at a distance of about three feet from the ring, intended to serve as a current detector. The ends of another coil were then connected to the terminals of a voltaic battery of ten cells. Calling the latter the primary circuit and the former the secondary circuit, it was found that, whenever the primary circuit was made or broken, a slight momentary deflection of the needle took place, indicating the presence of a transient current in the secondary circuit. Thus it was that Faraday made the fundamentally important discovery of the mutual induction of electric currents—a discovery which laid the foundation for all the modern applications of electricity requiring the supply of electric power in quantity. Faraday then proceeded to devise and construct further apparatus, in some of which the primary and secondary circuits could be moved relatively to each other. The results obtained may be summarised in the following statements:—

If part of the secondary circuit be near and parallel to the primary one, and the other parts, including the current indicator, at a greater distance, so that the indicator is not directly affected by the primary circuit, then, when the primary circuit is completed, a transient current in the opposite direction to that in the primary is produced in the secondary circuit. When the current

[1] A helix is the curve formed by the thread of a screw of uniform diameter throughout, and is often called a spiral; but the latter term is more properly applied to a curve in one plane, such as that formed by the mainspring of a watch.

in the primary circuit is broken, a current in the same direction as that in the primary circuit is produced in the secondary circuit.

The inductive effects are increased by bringing the parallel portions of the circuits close together, also by winding them into spiral or helical coils placed closed together, and still more by placing soft iron cores within the coils.

When the primary current is maintained constant and the conductor carrying it is at rest, the secondary current rapidly dies away; but if the two circuits are made to approach nearer together while the primary current is maintained constant, there will, during the approach, be a secondary current in the opposite direction to the primary one. If the two circuits are made to recede farther from each other, the primary current still being maintained constant, there will be a secondary current in the same direction as the primary one. Lenz showed, in 1834, that the direction of the secondary current produced by relative motions of the circuits is always such as to give rise to a mechanical force between the circuits tending to oppose the relative motion.

Similar results were obtained by causing relative motion to take place between the secondary circuit and a magnetic pole. Such results would necessarily follow from the equivalence of magnets and electric circuits established by Ampère.

The phenomena of electromagnetic induction may be very simply represented by means of the conception of lines of force when this is expressed in a quantitative form. We saw that in an electric field the lines of force always intersect the equipotential, or level, surfaces at right angles. In a magnetic field we may also speak of surfaces cutting the lines of force at right angles as level surfaces, since the magnetic force cannot have any component along such a surface. Suppose a small curve to be described on such a surface, and that the magnetic force across the small area enclosed by the curve be taken as the unit of magnetic force, then all the lines of force passing through the boundary of this small area will form a tube, which is called a unit tube of force. Any portion of the field of magnetic force will then be fully defined by these unit tubes, for the direction of the line of force at any point gives the direction of the force, while the number of unit tubes passing through unit area on the level surface through the point may be taken as a measure of the intensity of the force.

In every case of a current arising from electromagnetic induction it will be found that lines of magnetic force are cut by the conductor in which the current arises, and that as soon as this ceases the current will cease. It is easily seen that this cutting occurs when one circuit is moved relatively to the other; but it also takes place when an induced current arises from

the starting or stopping of a current, or, speaking more generally, from any variation of the current in a neighbouring conductor. If such a magnetic field is made to grow around the circuit in which the current is induced, or a field existing around it is made to die away, as long as any variation in the field causes lines of magnetic force to cut across the secondary circuit, an electromotive force, and therefore a current, will be produced in the latter.

The field of magnetic force surrounding an electric current may easily be exhibited by the aid of iron filings. Put a straight wire at right angles through a piece of cardboard supported horizontally, and pass a current through the wire: on lightly sprinkling iron filings on the cardboard they will be found to arrange themselves in a series of circles, having the hole as a common centre. The level surfaces, being perpendicular to the lines of force, will evidently consist of a series of planes intersecting each other in the axis of the straight wire. Since the lines of force are circles, it follows that if one pole of a magnet could be separated from the other it might be made to rotate continuously round a wire carrying an electric current. Although it is impossible to do this, the truth of the statement may be illustrated experimentally in a very simple manner. If a long, magnetised steel wire, fine enough to be easily flexible, is placed parallel and close to a wire carrying a current, the two poles will be found to rotate in opposite directions round the wire carrying the current, twisting the magnet round the wire. An electric current is therefore surrounded by what may be called "electric whirls."[1] When electrical actions are explained on Faraday's principles by actions going on in the surrounding medium, conductors carrying electric currents, and therefore also magnets, are found to give rise to actions of such a kind in this medium as might be imagined to arise from a rotation of portions of the medium. These rotations will give rise to electric whirls around any conductor moving across the lines of force, and will involve what is known as a current of electricity within the conductor.

Faraday found that the electric force which arises in a conductor at every point at which there is motion across the lines of magnetic force is always at right angles to the direction of motion and to the lines of magnetic force. The direction of the electric force along the straight line so determined he found to be given by the rule that if a man be supposed to swim in the

[1] These have usually been known as "magnetic whirls," but according to the representation of the electric field now generally accepted, and which is set forth on page 159, it is preferable to call them "electric whirls."

conductor so that he looks along the line of magnetic force in the positive direction, while the conductor moves towards his right hand, he will be swimming in the direction of the electric force, and therefore of the current, induced by the motion. Faraday found the amount of the resulting electromotive force in the conductor to be proportional to the number of unit tubes of magnetic force which the conductor cuts through per second, and therefore proportional to the intensity of the field and to the length and velocity of the conductor. If a closed circuit is moved across the lines of force, the passage of a unit tube from the inside to the outside of the area enclosed by the circuit will give rise to an electromotive force equal and opposite to that due to the passage of a unit tube from the outside to the inside. The phenomena of electromagnetic induction may therefore be summarised in the following statement:—

When an alteration takes place in the number of unit tubes of magnetic force passing in the same direction through the area enclosed by the secondary circuit, the electromotive force round the circuit will be measured by the rate of increase or decrease in the number of such tubes.

The reasoning by which this statement is arrived at would be invalidated if the portion of the electromagnetic field dealt with were divided into distinct regions by boundaries impenetrable to electromagnetic action (see p. 205).

When a current is started or stopped in any circuit it is clear that the resulting varying magnetic field will give rise to an electromotive force in the same circuit, the effect of which will be to oppose the growth, or the decay, as the case may be, of the current. This phenomenon is known as self-induction. It is easy to see that the self-induction of a straight wire of given dimensions would be greatly increased if it were wound into a coil, the effect of which would be to bring the different parts of the circuit much closer together, which would, by increasing their mutual induction, cause an increase in the self-induction of the whole.

The methods of producing electric currents hitherto considered only give rise to transient currents, and, although I have referred to Ampère's experiments with steady currents, I have not entered into the question of their production. We have seen that when an electromotive force gives rise to a current in a conductor, electricity is transferred from a place of higher to one of lower potential. If this had been effected by carrying a small charged body, such as a metal ball, from one place to the other, work would have been done on the ball by the electric forces. This work is utilised in the well-known experiment in which a small

ball is suspended by an insulating thread between an uninsulated bell and one which is kept charged by suspending it, by means of a metal wire, from the prime conductor of a frictional electrical machine. The ball is attracted to the charged bell, receives a similar charge, is repelled so as to strike the other bell, so that it loses its charge, and therefore it continues to oscillate between the two bells, and makes them ring continuously. When the electricity is transferred from one place to another by flow through a conductor no external work is done by the electric force, and therefore the principle of conservation of energy leads us to look for evidence of internal work being done in the conductor. This we find in the fact that the conductor becomes heated. In a simple metallic conductor all the internal work appears in the form of heat, and some of it in all cases. We cannot, therefore, expect to maintain a steady electric current without a continuous supply of energy.

As a result of work being done by an electric current traversing a conductor, it experiences a resistance to its passage. Ohm showed, in the year 1827, that this resistance has a definite value for every conductor, which varies with the nature of the conductor. The result of his investigations may be embodied in the following statement, which is known as Ohm's law:—The electromotive force acting between the extremities of any part of a circuit is the product of the strength of the current and the resistance of that part of the circuit. Ohm's law applies to all solid and liquid conductors, but we shall see that it does not apply to conduction through gases. If the portion of the circuit considered is not homogeneous, but contains transitions from one substance to another, or from one state of the same substance to another, internal electromotive forces may arise which will have to be taken into account.

The amount of energy converted into heat is measured by the product of the electromotive force into the quantity of electricity which passes. If W is the amount of heat energy generated in time t by a current C in a conductor of resistance R, due to an electromotive force E, then

$$W = ECt = C^2Rt \text{ (since } E = CR \text{ by Ohm's law).}$$

Faraday, in the course of his researches in electromagnetic induction, made a copper disc rotate between the poles of a horseshoe magnet, and found that a current was produced in a wire connecting the centre of the disc with its edge, and continued as long as the rotation was maintained. This continuous production of an electric current by the rotation of a conductor cutting the lines of force of a magnet is the essential principle of

the modern dynamo, which affords the most economical means of producing electric energy.

Other examples of the production of steady currents by means of mechanical energy are given by frictional and influence electrical machines. The former afford a convenient means of continuously producing electricity by friction, the latter by electric induction. Both of these enable us to produce extremely high electromotive force, especially the latter; but the resistance offered by the circuits is extremely high, so that the resulting current is so minute as to require a delicate galvanometer to indicate its presence.

Steady currents may also be obtained by means of primary and secondary batteries; the latter being commonly known as accumulators, though the term is not scientifically defensible. In both of these the current is maintained by the energy of chemical action. The essential difference between them is simply that the secondary battery is reversible, so that when it is exhausted it can be brought back to its original chemical condition by means of an electric current from an external source traversing it in the reverse direction to that which it produces, while a primary battery is not reversible, so that when it is exhausted fresh materials must be supplied. Before the invention of the dynamo, the primary battery formed the only practical source of electric current. Primary cells are commonly known as voltaic or galvanic cells, as they originated in the discoveries of Volta and Galvani, principally the former, towards the close of the eighteenth century. Several cells connected together form a voltaic or galvanic battery. As long as the current continues through a circuit containing either a primary or a secondary cell, chemical action will continue in the cell. When the circuit is broken, so that the flow of the current is arrested, the chemical action which had hitherto maintained the current is arrested, though in some kinds of cells local chemical action may continue, and cause the gradual destruction of the cell. Such local chemical action is always accompanied by electric currents completed within the cell itself.

All metals, whether in the liquid or solid state, allow electricity to pass through them without any resulting chemical change. Such a process is known as metallic conduction. The passage of electricity through all other conducting liquids is accompanied by chemical changes in the composition of the liquids. The process is called electrolytic conduction, and substances which conduct electricity in this manner when in the liquid state are known as electrolytes. Electrolytic conduction occurs in metallic salts and other chemical compounds when liquefied either by fusion or by solution in some solvent, usually water.

The chemical change accompanying the passage of a current of electricity through a liquid electrolyte is known as electrolysis, and is found to consist primarily in the decomposition of the substance into two distinct molecular groups which are known as ions. One of these groups is found to carry a positive electric charge, and is known as the "cathion." The other group carries a negative charge, and is known as the "anion." During the passage of the electric current through the liquid, the anion and the cathion travel in opposite directions until they reach the *electrodes*, or conductors by which the current enters and leaves the liquid respectively; the former electrode being known as the "anode," and the latter as the "cathode."

In some few simple cases the products of electrolysis, as liberated at the electrodes, are identical with the molecular groups forming the ions, but they are generally the result of the chemical action of the liberated ions on the solution. The vessel employed to contain the electrolyte for experimental work is known as a *voltameter*, or *voltametric cell*. When it is desired to maintain the electrolyte at a constant strength, an anode is employed of the metal of which the electrolyte is a salt. Otherwise, both electrodes should be of material unacted upon by the electrolyte, carbon or platinum being the most usual.

In the year 1803 Faraday summed up the results of his extensive researches into the facts of electrolysis in the two following general laws:—

1. The amount of chemical change which takes place in any electrolytic circuit is directly proportional to the quantity of electricity passing through the circuit.

2. For a given quantity of electricity passing through any circuit, the amounts of the different electrolytes decomposed in that circuit, and therefore the amounts of the ions set free, are proportional to the chemical equivalents[1] of the different electrolytes or ions.

Clerk-Maxwell, in the first edition of his classical treatise on electricity and magnetism, pointed out that Faraday's laws of electrolysis appeared to indicate a molecular nature for electricity, and suggested that such an assumption might serve provisionally as a convenient working hypothesis. He did not, however, regard such an hypothesis as one likely to be permanently retained when further advances should throw some light on the real nature of electricity.

H. von Helmholtz, in the course of the Faraday Lecture

[1] The chemical equivalent of any substance is the number representing the smallest mass of the substance which is capable of combining with unit mass of hydrogen, or of replacing it in any chemical combination.

delivered at the Royal Institution in 1881, argued that these laws necessarily led to the conclusion, that if the chemical elements were composed of atoms, then electricity, both positive and negative, must also be divided into definite elementary portions behaving like atoms of electricity; for the same definite quantity of either positive or negative electricity was found to move always with each univalent ion, or with every unit of affinity of a multivalent ion, accompanying it during all its motions through the interior of the electrolyte. This quantity he suggested might be called the electric charge of the atom. Dr Johnstone Stoney suggested that it should be called an electron, or natural electrical unit. Von Helmholtz pointed out, in the course of the same lecture, that decomposition would take place at the surface of the electrodes, provided the electromotive force were sufficient, and in that case the ions would deliver up their charges and become electrically neutral.

The cathode ray phenomena observed in vacuum tubes (see p. 107) led Varley to suggest, in 1871, that the ions really consisted of electrified particles projected from the cathode. Sir William Crookes's investigations confirmed this suggestion, and he, with wonderful intuition, designated the conditions of their existence as a *fourth state of matter*. These researches by Crookes led Sir J. J. Thomson to investigate the properties of a moving sphere carrying an electric charge, and in a paper of the most fundamental interest and importance, which appeared in the *Philosophical Magazine* in 1881, he showed that such a sphere would have a definite inertia, owing to its electric charge. This investigation may be considered as having laid the foundation for the electrical theory of matter, as it opened the way for an electrical explanation of the inertia, or mass, which is one of its most fundamental properties.

The probability that electrons might have a separate, independent existence, and possibly afford a basis for a general theory of physical phenomena, would appear to have occurred to several physicists very nearly simultaneously; the more prominent amongst them being Lorentz, Larmor, Voigt, and Drude.

Lorentz was the first to publish a definite electron theory, which he did in 1892, based mainly on the optical properties of media in motion relatively to each other, taking the astronomical phenomenon of the aberration of light (see Chapter VI.) as the starting-point. Larmor's starting-point was an attempt to apply MacCullagh's rotationally elastic ether (see Chapter IV.) to the representation of general electromagnetic theory, his attention having been directed to MacCullagh's theory by a paper published

in 1880 by Professor G. F. Fitzgerald, "On the Electromagnetic Theory of the Reflection and Refraction of Light." The first instalment of Larmor's theory was communicated to the Royal Society in 1893, and published in the *Philosophical Transactions* for 1894, with a supplement added in August 1894, in which the electron was definitely introduced in order to explain the relation of the observed forces between permanent magnets to the corresponding forces between circuits carrying electric currents, as this relation could not be accounted for on the ordinary vortex-atom theory.

CHAPTER II.

UNITS AND MEASUREMENT.

In the preceding chapter I have spoken of quantities of electricity and of electrical forces, and these terms will present no difficulty to any readers, as the ideas are perfectly familiar ones. In scientific studies, however, it is not sufficient to have a general apprehension of the meaning of the terms employed: it is further necessary, in order to form definite concepts which can be compared and reasoned about, to have the various quantities with which we are concerned expressed in terms of some definite system of units. Every expression of a quantity consists of two factors. One of these is the name of a fixed known quantity of the same kind as the quantity to be expressed, which is taken as the standard to which all other quantities of that kind are referred, and is called the unit. The other factor is the number expressing the ratio of the quantity in question to this unit. All quantities which are capable of expression in mechanical terms are capable of ultimate definition in terms of three fundamental units; and the three which are most conveniently taken as fundamental are the units of length, mass, and time. For scientific purposes, and, as far as electrical quantities are concerned, for commercial purposes as well, the fundamental units selected for the measurement of length, mass, and time, respectively, are the centimetre, the gramme, and the second.

Theoretically the centimetre is the thousand-millionth part of the length of a meridian of the earth measured from the pole to the equator, but practically it is a hundredth part of the standard metre preserved in Paris, when the standard is at the temperature of melting ice. The gramme is theoretically the mass of a cubic centimetre of distilled water at the temperature of melting ice and at the barometric pressure of 76 centimetres of mercury, but practically it is the thousandth part of the standard kilogramme preserved in Paris. The second is the mean solar second, which has been determined with great accuracy by astronomical observations. The system of derived physical units

based upon these three fundamental units is known as the C.G.S. system, these letters being the initial letters of the names of the fundamental units. The principal derived units of ordinary dynamics are those of velocity, acceleration, density, momentum, force, and work.

The velocity of a body moving uniformly is measured by the ratio of the distance described in any fixed time to the time taken to describe it; or, more generally, the velocity of a body at any moment is measured by the rate at which the length of its path is increasing relatively to the time. The unit of velocity is the velocity of a uniformly moving body which travels over a distance of a centimetre in a second.

The acceleration of a body at any moment is measured by the rate at which its velocity is increasing relatively to the time, and the unit of acceleration is the acceleration in which the velocity increases by a centimetre a second in a second.

The density of a substance is measured by the mass contained in the unit volume of the substance—that is to say, the mass of a cubic centimetre. The unit of density is therefore the density of a substance containing a gramme in each cubic centimetre.

The momentum of a body is measured by the product of its mass into its velocity, and the unit of momentum is therefore the momentum of a mass of one gramme moving with a velocity of a centimetre per second.

The force acting on a body is measured by the momentum produced per second, so that the unit of force is the force which produces unit momentum per second. This unit of force is called a *dyne*.

The work done by a force is measured by the product of the force into the displacement, in the direction of the force, of the point of a body at which the force is applied. The unit of work is therefore the work done by a force of a dyne acting through a centimetre in its own direction. This unit of work is called an *erg*.

The capacity of a system for performing work is called the energy of the system, and is measured by the amount of work which the system would be capable of performing by the expenditure of the whole of its energy.

The foundation for the science of electrical measurement was really laid by the investigations of Cavendish, about the middle of the eighteenth century, in which he demonstrated, by strict mathematical reasoning, that the experimental fact that a hollow globe carrying an electric charge does not part with any portion of its charge to a conductor in communication with, and enclosed within it, necessarily leads to the conclusion that the attraction between two small electrically charged bodies is inversely pro-

portional to the square of the distance between them. This fact was independently discovered, though not demonstrated with the same exactness, by Coulomb, who communicated his results to the French Academy in 1775. Cavendish had communicated some of his results to the Royal Society in 1771, but he never published his definite proof of this law; and the fact of his prior discovery of it was not generally known until the year 1879, when Cavendish's unpublished papers were edited by Clerk-Maxwell. Coulomb found, in his torsion-balance experiments, that the law held good for the attraction and repulsion between magnetic poles as well as for electrical actions. When Ohm discovered, as mentioned in the preceding chapter, the constancy of what we now call the resistance of a conductor, he did not take the further step of expressing the resistance of a circuit in terms of a standard unit, but confined himself to expressing the resistance of different portions of a circuit in terms of the resistance of a selected part of it. Pouillet, in 1837, took the very important step of expressing his resistance measurements in terms of the resistance of distilled mercury. The standard resistance employed by him was a column of mercury of a measured length contained in a U-shaped tube terminating in wide cups to permit of the necessary connections being made. It may be noted here that the reciprocal of the resistance of a conductor is called its conductivity, and that the resistance, and the conductivity, of a unit cube of a substance, measured between opposite faces, are called the specific resistance, and the specific conductivity, respectively.

In the meantime Gauss had published, in the year 1831, his important work on terrestrial magnetism, in which he showed that the strength of a magnetic pole could be expressed in terms of units based on the millimetre, the milligram, and the second. Weber, who had been associated with Gauss in his magnetic measurements, began in 1851 the development of a definite system of electrical measurements expressed in the terms of the same fundamental units. The importance of Weber's system was at once recognised by Lord Kelvin, and at his suggestion a Committee of the British Association was formed in 1861 for the consideration of standards of electrical resistance. The plan of work was subsequently extended so as to include the whole subject of electrical measurement. The co-operation of the leading British and foreign electricians was invited and obtained by this Committee, which included both purely scientific men and practical engineers; and the final conclusion at which they arrived was to adopt a series of practical units obtained by taking convenient multiples of theoretical units, based upon the centimetre, the gramme, and the second as fundamental ones. The

ultimate outcome of this Committee has been an international agreement arrived at by conferences between the electricians of all civilised countries, authorised by their respective Governments, as to the system of units to be employed in the measurement of electrical quantities. The most recent international conference on this subject took place in London in October 1908.

This uniformity has been rendered possible by the fact that the units so arrived at are what are called *absolute units*—that is to say, they are independent of the place as well as of the time at which the measurements are made. As an example of a unit which does not fulfil these conditions, I may take the weight of a pound, which has been, and still is, largely used by engineers as a measure of force. Where only moderate accuracy is required such a unit may do very well, but for really accurate work the place of measurement would have to be specified, since the weight of a pound is greater at the poles than at any other part of the earth's surface, and diminishes gradually as the equator is approached. It also depends upon any local circumstances which affect the value of gravity, such as the neighbourhood of mountains.

A system of electrical measurement might be based upon the definition of unit electric charge given in the Cavendish-Coulomb law referred to in Chapter I. According to this law, the force acting between two charges in a given dielectric medium is proportional to the product of the charges divided by the square of the distance between them. Therefore, unless a useless factor be introduced, the force may be defined as equal to this expression for some definite dielectric medium. The medium so chosen is air, in which the attraction is sensibly the same as *in vacuo*. From this it will follow that the unit charge will be that charge which repels a similar charge in air, at a distance of one centimetre, with a force of one dyne. A system of electrical measurement on this basis has actually been formulated, and is known as the electrostatic system. I have given the definition of the electrostatic unit of electric quantity because, as is shown later in this chapter, the number of electrostatic units of electric quantity contained in one electromagnetic unit represents a velocity, and this velocity is found to be the same as the velocity of light in the ether (see Chapter VIII., p. 157). The discovery of this fact formed the foundation-stone upon which Clerk-Maxwell based his theory of the electromagnetic character of the disturbances constituting light-waves. Except in relation to this question, it will be unnecessary for us to devote further consideration to the electrostatic system, since the electromagnetic system, founded on the definition of a unit magnetic pole, is the one universally employed.

Coulomb's experiments have shown that the force acting between two magnetic poles is of the same form as that between two electric charges. The unit magnetic pole may therefore be defined as that pole which repels a similar pole, at a distance of one centimetre, with a force of one dyne. Now the force exerted on a unit magnetic pole by a current whose distance from it is constant, that is to say, by a current flowing in a conductor in the form of a circular arc described with the pole as centre, was found by Ampère to be proportional to the length of the arc divided by the square of the radius of the circle. The current through the arc may therefore be determined in terms of a unit magnetic pole, by defining the force on a magnetic pole as being equal to the product of the length of the arc, the current through it, and the magnetic pole, divided by the square of the radius. The unit current will then be that current of which each centimetre exerts a force of one dyne on a unit magnetic pole at a distance of a centimetre. When measuring the actions between electric charges, and between magnetic poles and electric currents, all disturbing actions must of course be eliminated as far as possible; and the definitions only apply rigidly when such elimination is complete. The presence of air is not found to exercise sensible effects on the measurements as compared with those made *in vacuo* —that is to say, in free ether—and the presence of air need not therefore be considered as a disturbing agent. On the other hand, the intervention of a sheet of glass or mica between the electric charges when measuring their repulsion, or the neighbourhood of iron when measuring the actions between magnetic poles and currents, is found by experiment to exert considerable disturbing effects, and must therefore be avoided.

The unit electric charge, or unit of electric quantity, is defined in the electromagnetic system as the quantity conveyed by the unit current in one second.

We saw in Chapter I. that the potential at any point in an electric field is measured by the work done on a body carrying a unit quantity of electricity in moving it from that point to a point outside the electric field, or, as we may express it, to an infinite distance from all electrified bodies. The potential difference, or electromotive force, between two points in an electric field will therefore be measured by the work done in transferring a unit quantity of electricity from one point to the other. A restriction was pointed out in Chapter I., viz. that the body carrying the unit charge, and the unit selected, must be small enough not to disturb the existing electric field. It was necessary to make this restriction that the conditions should be such as not to alter the field, since otherwise the potential at a point in the

field would have no definite value. The reader will, however, see that it does not in any way affect the size of the unit selected for our system of measurement. If we wish to apply our definition to a field which would be changed by the actual motion of such a charge we may take as the moving charge a millionth of the unit, or as small a fraction as we please. Then the definition will take the form:—The potential difference, or electromotive force, between two points may be measured by a million times the work done in transferring a millionth of a unit quantity of electricity from one point to the other—which is equivalent to our former definition.

The unit electromotive force may therefore be defined as an electromotive force such that, if it be established between two points, an amount of work equal to one erg will be required to transfer the unit quantity of electricity from one to the other.

The letters P.D. are often used, for the sake of brevity, to represent the term potential difference; and the term electromotive force is similarly represented by the letters E.M.F. We saw in Chapter I. that in a purely electrostatic field the E.M.F. is identical with the P.D. In the general electromagnetic field part of the E.M.F. at any given point at a given instant will be due to the magnetic induction arising from the motion of conductors carrying electric currents (including magnets), and from the variation of the currents with time. In the absence of any general theory of the E.M.F.'s arising from chemical, thermoelectric, and other actions between the molecules of material bodies, we should have to treat these, when present, as disturbing effects to be separately considered and allowed for. On the electron theory, however, it is found that the E.M.F. may be considered as consisting of two parts only: one due to the magnetic induction, and the other derived from the potential of the distribution of fixed and moving electrons as existing at the instant considered, all these special E.M.F.'s being included in the latter part. We shall find, moreover, that when our analysis is extended so as to take account of the actions occurring within the magnetic molecules, the first part becomes included in the latter, so that the E.M.F. at any given instant may be derived entirely from the potential of the total distribution of fixed and moving electrons.

Having determined the units of current and of E.M.F., the unit of resistance can be defined, by means of Ohm's law, as the resistance of a conductor in which unit E.M.F. produces unit current.

The units obtained in this manner are not altogether of convenient magnitude for the expression of the electrical quantities which occur in practical work; some being too large, and others too small.

For this reason, the practical unit of electric current has been defined as one-tenth of the C.G.S. unit of current, and is called an ampere.

The unit of electromotive force is called a volt, after Volta, and is taken to be a hundred million C.G.S. units.

The unit of resistance, which is called an ohm, is then defined as the resistance of a conductor through which an electromotive force of one volt will maintain a current of one ampere. It contains a thousand million C.G.S. units.

The unit of power is called a watt, and is defined as the energy expended per second by an unvarying electric current of one ampere with a potential difference at its terminals (or under an electric pressure) of one volt. It is equivalent to ten million ergs per second.

The unit of quantity is called the coulomb, and is defined as the quantity of electricity carried in a second by a current of one ampere. It is equivalent to one-tenth of a C.G.S. unit.

The unit of capacity is called a farad, after Faraday, and is defined as the capacity of a condenser to which an E.M.F. of one volt will give a charge of one coulomb. As this unit is inconveniently large for most practical purposes, capacities are generally expressed in microfarads, or millionths of a farad.

The ohm, although a convenient unit for expressing the resistance of conductors, is inconveniently small for that of insulators, and their resistance is usually expressed in megohms, a megohm being equal to a million ohms. Through the influence of the British Association Committee, these units came into general use in Great Britain and her Colonies during the years 1870 and 1871. They were only, however, generally adopted after the meeting of the International Congress of Electricians which was held in Paris in 1881, when it was decided that they should be employed by electricians throughout the world.

Two further units have since been added, viz.:—

The unit of inductance, for the measurement of self and mutual induction, or *inductance*, is called a henry, and is defined as follows:—The self inductance of a circuit is said to be one henry when a current variation, at the rate of one ampere per second, in the circuit gives rise to an opposing electromotive force of one volt. The mutual inductance of two circuits is said to be one henry when a current variation, at the rate of one ampere per second, in one of the circuits gives rise to an electromotive force of one volt in the other circuit. The henry is equivalent to 10^9 C.G.S. units, viz. 10^9 centimetres, its dimensions being that of a length.

The unit of energy is called a joule, and is defined as the

amount of energy expended when a coulomb falls through a potential difference of a volt. It is equivalent to 10^7 ergs.

In the year 1894 the ohm, the ampere, and the volt were legally defined in Great Britain, and were at first known as legal units. Subsequently, however, they were adopted by other countries, and have therefore become known as international units. It was then decided to adopt the ohm and the ampere as the primary units; the former being determined in terms of a mercury standard, while the latter was determined by the electrolysis of a solution of nitrate of silver. The definitions were left practically unaltered at the recent meeting of the International Conference, and the definitions are given in full in the abstracts from their report which are appended to this chapter. The specifications, as well as the definitions, will be useful to the physical student for reference; and those whose interest in the subject is more of a general than of a detailed character will be enabled, if they so desire, to gain some idea from these specifications of the minute precautions adopted to secure accuracy in modern scientific measurements.

We have seen that the electrostatic system of units is founded on the definition of a unit charge of electricity as the charge which repels an equal charge at unit distance in air with unit force. The electromagnetic system, on the other hand, is founded on the definition of a unit magnetic pole as the pole which repels an equal pole at unit distance in air with unit force.

We might arrive at a third system of units based on the law of attraction and repulsion according to the inverse square of the distance by defining unit mass as the mass which attracts an equal mass at unit distance with unit force. This is called the gravitation system of units.

Now, these three systems of units are not consistent with each other, so that, for example, a quantity expressed in electrostatic units cannot be numerically compared with a similar quantity expressed in electromagnetic units, except by ascertaining the relation between the electrostatic and electromagnetic unit for the quantity in question. The reason for this inconsistency is very simple. All that we know experimentally in reference to the three systems is that the repulsion between two electric charges or between two magnetic poles, or the attraction between two masses, is proportional to the product of the two charges, magnetic poles, or masses, divided by the square of the distance separating them.

Let r be this distance in each case, F the force of repulsion or attraction, e an electric charge, m a magnetic pole, and M a mass;

then what the experimental results show is simply that, when A, B, and C are constants,

$$A\frac{e^2}{r^2}=B\frac{m^2}{r^2}=C\frac{M^2}{r^2}=F.$$

In order to express these relations in terms of defined physical constants, let us write

$$A=\frac{1}{K};\ B=\frac{1}{\mu};\ C=G.$$

Then we have

$$\frac{e^2}{Kr^2}=\frac{m^2}{\mu r^2}=\frac{GM^2}{r^2}=F.$$

Here K is the quantity defined as the dielectric constant, or the specific inductive capacity, of the medium, μ is the magnetic permeability, and G is the quantity known as the gravitation constant. The gravitational attraction does not, as far as we have been able to ascertain, depend in any way on the nature of the medium. It is therefore simply a number. The values of K and μ, on the other hand, are found to vary with the medium. The values of these quantities are considered in Chapter IX., and I may here mention that the value of G, as determined by Cavendish, is $6{\cdot}6\times10^8$; that μ is a quantity proportional to the density of the ether, being therefore measured, in C.G.S. units, in grammes per cubic centimetre, and that it probably exceeds the value 10^{12}, or a billion times the density of water; and that K is a quantity measured in cubic centimetres per erg, and that it is probably smaller than 10^{-32}.

The electrostatic system of units has been obtained by making $K=1$ for free ether, or in air, which is practically the same thing, so that the value of K thus obtained for any medium is not its absolute value for the medium, but the ratio of its value for the medium to its value for free ether.

The electromagnetic system of units has been obtained by making $\mu=1$ for free ether, or in air, which again is practically the same thing, so that the value of μ in electromagnetic units for any medium is the ratio of its absolute value for the medium to its absolute value for free ether.

Finally, the gravitational system of units, employed by Laplace and others in astronomical investigations, has been obtained by taking $G=1$.

If we know the "dimensions" of a quantity in terms of the fundamental units of length, time, and mass, we can always transform it from one system of units to another. To illustrate

this, take the very simple case of expressing a velocity of 60 miles per hour in feet per second.

Let v be the velocity and V the unit in the first case, and v' the velocity and V′ the unit in the second. Then if we write L and L′ and T and T′ for the corresponding units of length and time, we shall have $Vv = V'v'$, and therefore

$$\frac{v'}{v} = \frac{V}{V'} = \frac{\frac{L}{T}}{\frac{L'}{T'}} = \frac{LT'}{L'T} = \frac{1760 \times 3}{1} \times \frac{1}{60 \times 60} = 88:$$

so that 60 miles per hour is equivalent to 88 feet per second. The value of a unit in two systems may be compared in a similar manner, as, for example, the unit of electricity, taking L, T, and M as the units of length, time, and mass. Then, both in the electrostatic and electromagnetic systems, L = 1 centimetre, T = 1 second, M = 1 gramme. Now, in the electrostatic system, the dimensions of e^2 are:

$$\text{Force} \times L^2 = \text{mass} \times \text{acceleration} \times L^2 = \frac{ML}{T^2} \times L^2;$$

so that the dimensions of e are

$$M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}.$$

In the electromagnetic system we have

$$\text{Force on magnetic pole} = \frac{\text{length of arc} \times \text{current} \times \text{magnetic pole}}{(\text{distance})^2}.$$

Therefore

$$\text{Current} = \frac{\text{force} \times (\text{distance})^2}{\text{length of arc} \times \text{magnetic pole}} = \frac{MLT^{-2} \times L^2}{L \times M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}} = L^{\frac{1}{2}}M^{\frac{1}{2}}T^{-1}.$$

But quantity e = current × time, so that the dimensions of e are

$$L^{\frac{1}{2}}M^{\frac{1}{2}}.$$

Therefore

$$\frac{\text{Electrostatic unit of electric quantity}}{\text{Electromagnetic unit of electric quantity}} = \frac{M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}}{L^{\frac{1}{2}}M^{\frac{1}{2}}}$$
$$= LT^{-1} = \text{velocity}.$$

When the value of this velocity is determined by measuring the ratio electrically, it is found to lead to the same result, within the limit of observational errors, as the measurement of the velocity of light. It was this agreement which first led Maxwell

to formulate his electromagnetic theory of light, since this velocity represents the speed of propagation of an electromagnetic disturbance through the ether.

Maxwell's equations of electromagnetic wave-propagation through the ether show that if c is this velocity, which we shall call the constant of radiation, then

$$c = \frac{1}{\sqrt{K\mu}}.$$

The approximate value of c in C.G.S. units is

3×10^{10} centimetres per second.

We may show that the product $K\mu$ has the dimensions of the reciprocal of a velocity squared, as is required by the above equation, in a very simple manner indicated by Sir Oliver Lodge.[1]

Let L, A, v, F, T, Q, m represent units of length, area, velocity, force, time, electric quantity, and magnetic pole respectively.

Then, by the definition of Q and m, we have

$$Q = L\sqrt{KF} \quad \text{and} \quad m = L\sqrt{\mu F}.$$

Now we shall see in Chapter VIII. (p. 139) that the relation between an electric circuit and the magnetic moment of the equivalent magnet is given by the equation

$$ml = \mu AC, \quad \text{or} \quad m = \mu v Q. \quad \text{Therefore}$$

$$\sqrt{\frac{\mu}{K}} = \frac{m}{Q} = \mu v$$

or

$$\mu K = \frac{1}{v^2} = \frac{\text{density}}{\text{elasticity}}.$$

We shall see in Chapter VIII. that Maxwell takes $\frac{4\pi}{K}$ as representing the electric elasticity. If K_0, μ_0 are the values of K and μ for the ether, and the elasticity of the ether is taken to be $\frac{4\pi}{K_0}$, then (the value c being at the same time taken for v) the above equations indicate that the density ρ of the ether must have the value $4\pi\mu_0$.

The reader will have no difficulty in making out for himself, on the lines of the above example, the ratio of any electrostatic unit to its corresponding electromagnetic unit; he will then find that these ratios may all be expressed in terms of v, viz.:—

In the cases of electric quantity, current in a circuit, magnetic

[1] See *Modern Views of Electricity*, Appendix.

potential, electric displacement, magnetic force, current at a point, there are v electrostatic units in one electromagnetic unit.

In the cases of magnetic pole, electrokinetic momentum of a circuit, electromotive force round a circuit, magnetic induction, electromotive force at a point, vector potential, there are $1/v$ electrostatic units in one electromagnetic unit.

In the cases of electrostatic capacity, specific inductive capacity, conductivity, there are v^2 electrostatic units in one electromagnetic unit.

In the cases of electromagnetic capacity or coefficient of self-induction, magnetic permeability, resistance, there are $1/v^2$ electrostatic units in an electromagnetic unit.

Extracts from the Report of the International Conference on Electrical Units and Standards, London, October 1908.

The Conference elected a Technical Committee to draft specifications and to consider any matter which might be referred to the Committee and to report to the Conference. The Conference and its Technical Committee each held five sittings. As a result of its deliberation the Conference adopted the resolutions and specifications attached to this report and set out in schedule B, and requested the delegates to lay them before their respective Governments with a view to obtaining uniformity in the legislation with regard to Electrical Units and Standards.

The Conference recommend the use of the Weston Normal Cell as a convenient method of measuring both electromotive force and current when set up under the conditions specified in schedule C.

In cases in which it is not desired to set up the Standards provided in the resolutions, schedule B, the Conference recommends the following as working methods for the realisation of the International Ohm, the Ampere, and the Volt:—

1. *For the International Ohm*—

The use of copies, constructed of suitable material and of suitable form and verified from time to time, of the International Ohm, its multiples and submultiples.

2. *For the International Ampere*—

(*a*) The measurement of current by the aid of a current balance standardised by comparison with a silver voltameter;

(*b*) The use of a Weston Normal Cell whose electromotive force has been determined in terms of the International Ohm and

International Ampere, and of a resistance of known value in International Ohms.

3. *For the International Volt*—

(*a*) A comparison with the difference of electrical potential between the ends of a coil of resistance of known value in International Ohms, when carrying a current of known value in International Amperes; or

(*b*) The use of a Weston Normal Cell whose electromotive force has been determined in terms of the International Ohm and the International Ampere.

SCHEDULE B.

Resolutions.

I. The Conference agrees that as heretofore the magnitudes of the fundamental electric units shall be determined on the electromagnetic system of measurement with reference to the centimetre as the unit of length, the gramme as the unit of mass, and the second as the unit of time.

These fundamental units are (1) the Ohm, the unit of electric resistance which has the value of 1,000,000,000 in terms of the centimetre and second; (2) the Ampere, the unit of electric current which has the value of one-tenth (0·1) in terms of the centimetre, gramme, and second; (3) the Volt, the unit of electromotive force which has the value 100,000,000 in terms of the centimetre, the gramme, and the second; (4) the Watt, the unit of power which has the value 10,000,000 in terms of the centimetre, the gramme, and the second.

II. As a system of units representing the above and sufficiently near to them to be adopted for the purpose of electrical measurements and as a basis for legislation, the Conference recommends the adoption of the International Ohm, the International Ampere, and the International Volt defined according to the following definitions.

III. The Ohm is the first Primary Unit.

IV. The International Ohm is defined as the resistance of a specified column of mercury.

V. The International Ohm is the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14·4521 grammes in mass, of a constant cross-sectional area and of a length of 106·300 centimetres.

To determine the resistance of a column of mercury in terms

of the International Ohm, the procedure to be followed shall be that set out in specification I. attached to these Resolutions.

VI. The Ampere is the second Primary Unit.

VII. The International Ampere is the unvarying electric current which, when passed through a solution of nitrate of silver in water, in accordance with the specification II. attached to these resolutions, deposits silver at the rate of 0·00111800 of a gramme per second.

VIII. The International Volt is the electrical pressure which, when steadily applied to a conductor whose resistance is one International Ohm, will produce a current of one International Ampere.

IX. The International Watt is the energy expended per second by an unvarying electric current of one International Ampere under an electric pressure of one International Volt.

Specification I.

SPECIFICATION RELATING TO MERCURY STANDARDS OF RESISTANCE.

The glass tubes used for mercury standards of resistance must be made of a glass such that the dimensions may remain as constant as possible. The tubes must be well annealed and straight. The bore must be as nearly as possible uniform and circular, and the area of cross-section of the bore must be approximately one square millimetre. The mercury must have a resistance of approximately one ohm.

Each of the tubes must be accurately calibrated. The correction to be applied to allow for the area of the cross-section of the bore not being exactly the same at all parts of the tube must not exceed 5 parts in 10,000.

The mercury filling the tube must be considered as bounded by plane surfaces placed in contact with the ends of the tube.

The length of the axis of the tube, the mass of mercury the tube contains, and the electrical resistance of the mercury are to be determined at a temperature as near to 0° C. as possible. The measurements are to be corrected to 0° C.

For the purpose of the electrical measurements, end vessels carrying connections for the current and potential terminals are to be fitted on to the tube. These end vessels are to be spherical in shape (of a diameter of approximately four centimetres) and should have cylindrical pieces attached to make connections with the tubes. The outside edge of each end of the tube is to be coincident with the inner surface of the corresponding spherical

end vessel. The leads which make contact with the mercury are to be of thin platinum wire fused into glass. The point of entry of the current lead and the end of the tube are to be at opposite ends of a diameter of the bulb; the potential lead is to be midway between these two points. All the leads must be so thin that no error in the resistance is introduced through conduction of heat to the mercury. The filling of the tube with mercury for the purpose of the resistance measurements must be carried out under the same conditions as the filling for the determination of the mass.

The resistance which has to be added to the resistance of the tube to allow for the effect of the end vessels is to be calculated by the formula—

$$A = \frac{0{\cdot}80}{1063\pi}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \text{ ohm},$$

where r_1 and r_2 are the radii in millimetres of the end sections of the bore of the tube.

The mean of the calculated resistances of at least five tubes shall be taken to determine the value of the unit of resistance.

For the purpose of the comparison of resistances with a mercury tube the measurements shall be made with at least three separate fillings of the tube.

Specification II.

SPECIFICATION RELATING TO THE DEPOSITION OF SILVER.

The electrolyte shall consist of a solution of from 15 to 20 parts by weight of silver nitrate in 100 parts of distilled water. The solution must only be used once, and only for so long that not more than 30 per cent. of the silver in the solution is deposited.

The anode shall be of silver, and the kathode of platinum. The current density at the anode shall not exceed 1/5 ampere per square centimetre and at the kathode 1/50 ampere per square centimetre.

Not less than 100 cubic centimetres of electrolyte shall be used in a voltameter.

Care must be taken that no particles which may become mechanically detached from the anode shall reach the kathode.

Before weighing, any traces of solution adhering to the kathode must be removed, and the kathode dried.

SCHEDULE C.

Weston Normal Cell.

The Weston Normal Cell may be conveniently employed as a standard of electric pressure for the measurement both of E.M.F. and of current, and when set up in accordance with the following specification, may be taken, provisionally, as having, at a temperature of 20° C., an E.M.F. of 1·0184 volts.

Specification relating to the Weston Normal Cell.

The Weston Normal Cell is a voltaic cell which has a saturated aqueous solution of cadmium sulphate ($CdSO_4$ 8/3 H_2O) as its electrolyte.

The electrolyte must be neutral to Congo Red.

The positive electrode of the cell is mercury.

The negative electrode of the cell is cadmium amalgam consisting of 12·5 parts by weight of cadmium in 100 parts of amalgam.

The depolariser, which is placed in contact with the positive electrode, is a paste made by mixing mercurous sulphate with powdered crystals of cadmium sulphate and a saturated aqueous solution of cadmium sulphate.

The different methods of preparing the mercurous sulphate paste are described in notes appended to the report of the conference. One of the methods there specified must be carried out.

For setting up the cell, the H form is the most suitable. The leads passing through the glass to the electrodes must be of platinum wire, which must not be allowed to come into contact with the electrolyte. The amalgam is placed in one limb, the mercury in the other.

The depolariser is placed above the mercury and a layer of cadmium sulphate crystals is introduced into each limb. The entire cell is filled with a saturated solution of cadmium sulphate and then hermetically sealed.

The following formula is recommended for the E.M.F. of the cell in terms of the temperature between the limits 0° C. and 40° C.:—

$$E_t = E_{20} - 0{\cdot}0000406(t - 20^\circ) - 0{\cdot}00000095(t - 20^\circ)^2 + 0{\cdot}00000001(t - 20)^3.$$

CHAPTER III.

MEANING AND POSSIBILITY OF A MECHANICAL THEORY OF ELECTRICITY.

THE simplest branch of physics, and the one which was first developed into a connected scheme, is that concerned with the motions and the equilibrium of bodies, and known as mechanics. The earliest investigations of which we have any records were those of the ancient Greeks, who concerned themselves exclusively with the conditions of equilibrium, originally in connection with such simple devices as levers, pulleys, and inclined planes. There appear to have been no attempts to inquire into the conditions of motion prior to the researches of Galileo in the sixteenth and seventeenth centuries. In the course of his investigations into the laws of falling bodies he introduced the idea of acceleration, and with it the modern conception of force, hitherto known only in the form of pressure. In this way he laid the foundation for the dynamics of a single body, and soon afterwards Huygens solved the first problems of systems of bodies. The latter was closely followed by Newton, who published his great treatise on *The Mathematical Principles of Natural Philosophy* in the year 1686. In this work he set forth his discovery of universal gravitation, which so greatly extended the scope of mechanical theory, and so far completed the formal enunciation of mechanical principles, that the advance since his time has consisted almost entirely of formal mathematical development on the broad basis of Newton's laws of motion.

The form in which the law of gravitation was enunciated by Newton led to the expression of the conditions of equilibrium and of motion of systems of bodies in terms of forces acting at a distance between their component particles, and to a corresponding development in mathematical methods for the investigation of these systems of central forces. Newton himself was never satisfied with systems of action at a distance; but though physicists have evolved more satisfying and complete representations for the greater portion of the wide range of physical phenomena, the most promising attempts to account for gravitation are still in the

hypothetical stage, and even these hypotheses involve, as we shall see in Chapter XXIII., what can only be designated as extra-physical concepts. For more than a century after the time of Newton the central force theory reigned supreme, and it was only to be expected that, when the first attempt was made to frame a theory of electric and magnetic phenomena, it should be constructed on the lines of the theory of central forces. Accordingly, it was assumed that electric attractions and repulsions were due to central forces similar to gravitational forces, and acting between the particles of "electric matter," which was assumed to possess the properties requisite to account for the phenomena observed. Magnetic phenomena were ascribed in a similar manner to distributions of "magnetic matter." In the hands of Cavendish, Coulomb, Poisson, and others, a complete scheme was developed which embraced all the known electrostatic phenomena and a considerable range of magnetic phenomena. Ampère formulated a theory of the interaction of electric currents, based on the assumption of repulsion and attraction, following the same law as the attraction of gravitation, between every pair of elements of a system of currents. This theory gave very complicated and artificial-looking expressions for the mutual actions of currents in unclosed circuits; but the actions of closed circuits, which alone could be experimentally verified, were in complete agreement with the results of the theory. With the further extensions which it received at the hands of Weber, Neumann, and others, this hypothesis gave a satisfactory representation of a very wide range of phenomena, including the explanation of magnetic effects in terms of molecular electric currents. Lord Kelvin and von Helmholtz have shown that Faraday's great discovery of electromagnetic induction might have been anticipated from this theory, and Kirchhoff has proved that it would account for the propagation of electric waves of short periods along a metallic wire with a velocity nearly equal to that of light. In fact, as developed by Weber, this theory accounted for the whole range of phenomena as then known, and would have accounted for others not then discovered. Weber's theory was adversely criticised by von Helmholtz on general dynamical considerations, but his arguments really only showed that the scheme led to impossible results in the case of velocities, or vibration frequencies, approaching those of light—conditions of which there was no experimental evidence prior to Hertz's researches on electric waves in 1888, many years later. That a theory so comprehensive should have to be discarded in favour of one still more far-reaching is a striking fact, but by no means an isolated one. The path of progress in physical science is strewn with theories which have

served their purpose, and served it well, by leading the way to their own replacement by more perfect ones, which will in all probability have to give way in their turn. I hope to show presently that this result is what is to be expected from an essential characteristic of mechanical representations of physical phenomena, and that in fact it could not well be otherwise.

Sir Joseph Larmor, in his treatise *Æther and Matter*, draws attention to a letter written by Weber to Gauss in 1845, in which the former refers to his unsuccessful attempts to account for the forces between moving electric particles on the basis, not of instantaneously propagated actions between them, as hitherto assumed, but of actions propagated at a finite speed, as in the case of light. He had the instinctive conviction that the desired result would not be attained until a consistent representation had been formed of the manner in which the propagation can take place. As Larmor observes, this consistent representation could be nothing else than a dynamical theory of the luminiferous medium, the ether. We shall see in Chapter IV. that such a scheme had been arrived at on the basis of purely optical theory by MacCullagh in 1839, and his solution was identical in form with the one arrived at by Maxwell some twenty years later, from the consideration of a much wider range of phenomena. In the meantime, the magnificent series of discoveries made experimentally by Faraday, in following out his ideas of the continuous transmission of physical actions, were providing materials for a consistent electrical theory based upon these considerations. Faraday, indeed, accomplished much more than the accumulation of the material, and the commanding importance of his services in that respect must not be allowed to obscure the value of his constructional work in laying the essential foundations of a mechanical theory of electricity based upon the idea of the continuous transmission of all physical actions. The first aspect of Faraday's work was fully appreciated by his contemporaries, but the importance of its second aspect was first grasped by the genius of Clerk-Maxwell. Faraday's theory had the complete consistency of a theory developed on rigidly mathematical lines; but Faraday was not a mathematician, and consequently his investigations were not cast in the mathematical mould. The images employed by him were for that reason far more difficult to grasp, and it was only when Maxwell had developed them in mathematical form that their import was fully understood. The matter is of such interest and importance that I need make no apology for reproducing Maxwell's account of it in his own words, as given in the preface to his treatise on *Electricity and Magnetism*, published in 1873. Here he says:—

"Before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's *Experimental Researches on Electricity*. I was aware that there was supposed to be a difference between Faraday's way of conceiving phenomena and that of the mathematicians, so that neither he nor they were satisfied with each other's language. I had also the conviction that this discrepancy did not arise from either party being wrong. I was first convinced of this by Sir William Thomson, to whose advice and assistance, as well as to his published papers, I owe most of what I have learned on the subject.

"As I proceeded with the study of Faraday, I perceived that his method of conceiving the phenomena was also a mathematical one, though not exhibited in the conventional form of mathematical symbols. I also found that these methods were capable of being expressed in the ordinary mathematical forms, and thus compared with those of the professed mathematicians.

"For instance, Faraday, in his mind's eye, saw lines of force traversing all space where the mathematicians saw centres of force attracting at a distance: Faraday saw a medium where they saw nothing but distance: Faraday sought the seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids.

"When I had translated what I considered to be Faraday's ideas into a mathematical form, I found that in general the results of the two methods coincided, so that the same phenomena were accounted for, and the same laws of action deduced by both methods, but that Faraday's methods resembled those in which we begin with the whole and arrive at the parts by analysis, while the ordinary mathematical methods were founded on the principle of beginning with the parts and building up the whole by synthesis.

"I also found that several of the most fertile methods of research discovered by the mathematicians could be expressed much better in terms of ideas derived from Faraday than in their original form.

"The whole theory, for instance, of the potential, considered as a quantity which satisfies certain partial differential equations, belongs essentially to the method which I have called that of Faraday. According to the other method, the potential, if it is to be considered at all, must be regarded as the result of a summation of the electrified particles divided each by its distance from a given point. Hence many of the mathematical discoveries of Laplace, Poisson, Green, and Gauss find their proper place in this treatise, and their appropriate expression in terms of conceptions mainly derived from Faraday."

Maxwell did not himself develop a complete mechanical theory of electricity based on Faraday's ideas of the continuous transmission of every kind of physical action. This was not a work to be accomplished by one man, however great his genius. But Maxwell made a step in advance of the most fundamental importance by developing a train of argument which, subject to a single important correction, subsequently supplied by Larmor, definitely established the possibility of such a mechanical theory.

This single weak point in the argument was the assumption that the specification of an electric current in terms of the continuous series of current elements employed by Ampère was physically adequate. Larmor showed that this was not the case, and completed the demonstration by substituting the discrete elements of the electron theory, and showing that this specification was an adequate one.

In order to obtain a clear conception of the meaning of this step, and of the manner in which it was made, we must consider the methods by which mechanical generalisations, or laws, have been arrived at. A very good example is afforded by the procedure adopted by Archimedes in the third century B.C., in his studies of pressure in liquids. This consisted essentially in regarding the pressure due to a quantity of liquid as compounded of the pressures arising from a number of similar smaller portions into which he imagined the whole quantity to be divided. He thus resolved a comparatively complex phenomenon into a number of simple elements which were all of the same kind, and in so doing he made mathematical physics possible. It is the province of mathematics to deal with the combinations of quantities of the same kind, and therefore, when a physical problem can be expressed in terms of such elements, it is not in general necessary to reason it out completely from its elements in steps, each consisting of reasoning specially devised for application to the problem before us. Much of this can be obviated by the use of mathematical machinery which may perhaps have been elaborated for the solution of a very different physical problem, or possibly by workers interested only in the development of the machinery itself, without thought or care of its possible applicability to the extension of our knowledge of physical phenomena. It will be seen from this that the application of mathematical reasoning to the results of observations can never lead to any conclusions which were not implicitly contained in the observations themselves. What it does for us is to demonstrate the existence of many consequences, which, except for mathematical assistance, would have remained hidden from our view, buried, in fact, beneath the complexity of the observed results. As a second example we may take Galileo's investigation of the relation between the velocity acquired by a body falling from rest, and the time of fall. The method he adopted was to assume what appeared to him a probable relation, and then to compare the results of the hypothesis with those obtained by observation. Assuming that the velocity increased by equal amounts in equal times, the question arose as to the best method of testing the correctness of the assumption. The direct measurement of the velocities would

be a difficult matter, and it would be much easier to measure the distance fallen through. But if the assumption made were correct, these distances would be proportional to the squares of the times; and conversely, if the latter relation held, the validity of the assumption would be demonstrated. Experiment showed this to be the relation connecting the times and the distances, and therefore the truth of the assumption was established. The derivation of the relation between the distance and the time from that between the velocity and the time can be effected either by a simple algebraical summation, or by an easy geometrical construction, on the basis of a division of the time of fall into a number of equal parts. Both the distance fallen through and the velocity are, in the case considered, what are called functions of the time, viz. they undergo changes in value as the time changes. The relation of the velocity to the distance is a typical example of a relationship which is of primary importance in the mathematical treatment of physical problems. The ratio of the distance traversed by a moving particle to the time occupied in traversing it is defined to be the average, or mean, velocity of the particle during that interval of time. If the interval of time taken be so small that there is no sensible change in the velocity during that interval, the ratio will represent the velocity at an instant included in the interval, and may be defined as the rate of change, at the given instant, of the distance traversed, relative to the time. The consideration of quantitative relationships of this kind had not previously arisen, and new mathematical methods were devised by Newton and Leibnitz for their treatment. The velocity is defined from this point of view as the first derivative of the distance with respect to the time. The acceleration, first definitely introduced by Newton, bears the same relation to the velocity that the velocity bears to the distance traversed. That is to say, the acceleration is the rate of change of the velocity with respect to the time, and mathematically it is therefore the first derivative of the velocity with respect to the time. It is therefore also said to be the second derivative of the distance described with respect to the time. The time in such cases is said to be the independent variable.

The principles enumerated by Newton made it theoretically possible to determine the future movements of any system of bodies consisting of specified gravitating masses, occupying specified positions, and having specified velocities, at a given instant. Suppose, for the sake of simplicity, that all the bodies are so small compared with their relative distances that no sensible error will be introduced by treating them as if the position of each could be considered as fixed by that of a single

point in it, at which its whole mass may be supposed to be concentrated. Then, if every pair of points be joined by a straight line, along every such line joining two particles there will arise, according to Newton's law, a relative acceleration of the two particles, which will be determined by their masses, and the distance separating them at the instant of time at which the system is considered. The result of these accelerations will be to superpose on the initial velocity of each particle a new velocity in a different direction. Consider the system again after the lapse of an interval of time so short that the effect of the changes in the accelerations due to the altered positions of the particles may be neglected. The displacement of each particle may then be obtained as the resultant of the displacement due to its initial velocity and of the accelerations due to each of the other particles of the system. The new accelerations may then be determined in the new relative positions of the particles, and used to determine the changes in these positions after a second small interval of time. In the case of a small number of particles, all in the same plane, the successive positions can be represented without any difficulty by constructing successive diagrams on a drawing-board; but the process would be a very tedious one, and, except in simple cases, would be impossible to carry out practically.

Although its nature would remain unchanged, the problem would become still more complex if the particles were aggregated into clusters. It was in order to deal with these difficulties that the new mathematical methods, now known as differentiation and integration, were devised independently by Newton and Leibnitz. The motion of such a system will be fully determined if we can express the positions of the masses at any time in terms of the initial positions and velocities and of the time which has elapsed since the initial state. The first step is to fix the positions of the particles by means of what are called co-ordinates. A familiar example is afforded by the determination of the position of a point on the earth's surface by means of its latitude and longitude. A convenient method very generally used to determine the position of a point in space is to take as its co-ordinates its distances from a set of three planes at right angles to each other. The distances between the particles can also be expressed in terms of these co-ordinates, and so can the velocities and accelerations, viz. as the first and second derivatives of the distances with respect to the time. An equation, or set of equations, can then be formed, expressing the connection between the co-ordinates and their first and second derivatives with respect to the time. In the case of the system of co-ordinates referred to above, the distances between the particles are replaced by their components parallel to the three

lines of intersection of the planes of reference, and three sets of equations are obtained, each expressing the relations between distances measured parallel to one of these lines of intersection, or axes of co-ordinates, and the first and second rates of change of these distances with respect to the time, representing the velocities and accelerations parallel to the same axes. These equations are called differential equations, and, because they do not contain any higher derivatives than the second, they are said to be of the second order. The possibility of expressing the conditions of motion by differential equations is a mere mathematical consequence of the fact that the successive states of such a system are derived from each other by continuous changes, and not by sudden jumps, so that the state of such a system can always be determined from the condition immediately preceding it in time, without taking into account what may have been the conditions at more remote periods. The restriction of the order of the equations to the second is due to the comparatively simple nature of the experimental laws which have been found adequate to the specification of the motions of systems of masses.

Any group of phenomena of which the sequence of changes in time can be shown to be such that the relations between them at any instant can be determined from a knowledge of the relations existing at an instant immediately preceding it may be said to be capable of mechanical representation, and this condition may be taken as defining the meaning of a mechanical representation in its widest possible sense.

As a matter of practical convenience, it is found advisable to restrict this widest meaning of the term mechanical representation by the further condition that it shall conform to the general principles of dynamics as contained explicitly or implicitly in Newton's laws of motion, and it is in this restricted sense only that I shall employ the term. The resulting loss of generality is more apparent than real, for experience has invariably shown that representations which did not come within the restricted meaning of the term have had to be rejected, owing to their want of conformity with facts of observation. The effect of the restriction is, therefore, merely to prevent waste of time and effort in formulating unfruitful hypotheses. Before proceeding to consider the simplest and most convenient expression of Newton's principles for the purpose in view, I will make a slight digression to consider a point which, although not strictly involved in our present argument, may have occurred to some of my readers as a practical difficulty in applying Newton's principles to the determination of the successive states of a material system. I have already pointed out that this problem is made more complex when the particles,

instead of being separated in space, are aggregated into clusters, so as to form what we call material bodies. The complexity involved in passing from separate particles to a continuous distribution of matter in space is exactly analogous to that met with in passing from a series of separate successive states to a continuous variation in time, and similar mathematical methods are available for overcoming the resulting difficulties. When the system varies continuously with the time, we have to express the velocity and acceleration as derivatives of the time as an independent variable. Instead of proceeding from point to point in time, we now have to proceed from point to point in space, and therefore, if no further difference were involved, we should simply have to consider the rates of change in the velocity and acceleration due to change in position, instead of those due to change in time. That is to say, speaking mathematically, we should have to express these quantities as derivatives of the co-ordinates instead of expressing them as derivatives of the time. There is, however, a further difference. In the case of the succession in time, we are always dealing with the same particles, while, in changing from one position to another in space, this is not the case. Owing, however, to the homogeneity of matter, the successive sets of particles encountered during the spatial changes of the system usually resemble one another so closely that no appreciable error is introduced by treating them as if they were identical. This assumption is not always allowable. Thus, if we were considering the motions of an iron bowl filled with water, we should arrive at incorrect results if we treated the whole mass as consisting of similar matter. We should then have to consider the iron and the water separately, but the continuous methods of mathematical treatment would still be separately applicable within the boundaries of the iron and the water respectively.

It does not follow from the theoretical sufficiency of the Newtonian principles for the treatment of any possible mechanical problem that they will necessarily enable us to solve any and every such problem that may suggest itself. We have seen that, in order to apply them, it is necessary for us to enter into very considerable detail in the specification of the problem, and this is not always possible, and in the case of problems in molecular and atomic physics it is almost invariably impossible. Newton's methods were purely geometrical, and therefore suitable only for what is known as the *synthetic* method, in which the solution of a problem is sought by making deductions from given assumptions, and then comparing these deductions with the observed facts. Moreover, in following out this system, no general method of procedure can be devised once for all. Each problem has to be

considered separately, and will generally involve the use of special, and often by no means obvious, artifices. In the more general *analytical* method, which has been rendered possible by the application of algebraic methods to geometry, we endeavour to determine the conditions of motion of a system directly from data in the form of observed facts. The first important advance in this direction was made by d'Alembert in his enunciation, in 1743, of a general principle which enabled results arrived at in the investigation of problems of equilibrium to be utilised in dealing with questions relating to the mutual actions of systems of particles having fixed connections between them. This step, however, did not enable the consideration of details to be dispensed with. It merely facilitated investigations by enabling previously acquired knowledge of these details to be utilised in many cases where, otherwise, it would have been necessary to work them out independently.

An unsuccessful attempt was made by Maupertuis in 1747 to formulate a single general principle, in the form of a law of "least action," of the change from one state of a system to another, and which he propounded as a principle eminently according with the wisdom of the Creator. No definite expression of such a principle was ever obtained by Maupertuis, but his attempts to obtain such an expression led Euler to the formulation of a definite simple expression which, in the case of a particle moving from one position to a neighbouring one, will have a smaller value for the actual path than for any other possible closely adjacent one, always provided that the velocity depends on no other variable quantities than those which determine the position of the particle.[1] The view taken by Euler was that the purposes of natural phenomena afford as satisfactory a basis for their explanation as do their causes. He therefore assumed the *a priori* probability of the existence of a maximum or minimum condition in the case of any natural phenomena. He did not, however, expect to be able to ascertain the character of such a maximum or minimum by mere metaphysical reasoning, but sought, in the course of the ordinary procedure for the solution of mechanical problems, to find an expression which should in all cases be a maximum or a minimum. It was subsequently pointed out by Jacobi that, although this condition is very generally fulfilled, it is not

[1] *Mathematical Note.*—Euler's principle is that, when v is a function only of the co-ordinates of the particle, the expression $\int v ds$ will be a minimum for the actual path of the particle, where ds is an element of the path, and v is the corresponding velocity. The path can therefore be determined by expressing mathematically the condition that $\int v ds$ should be a minimum.

universally so. Euler had already, in 1736, laid the foundations for the analytical treatment of mechanical problems, and further important advances were made by MacLaurin in a treatise published in 1742. On the foundations laid by these two eminent mathematicians there appeared in 1788 an edifice of imposing grandeur and beauty, Lagrange's system of analytical mechanics. Lagrange had attempted in a youthful work to found mechanics on Euler's principle of "least action," and extended this principle to systems of masses;[1] but in the *Mécanique Analytique* he expressly declares his intention of completely discarding all metaphysical speculation as a basis for dynamics. His system of statics was based on what is now known as the principle of work, but was then known as the principle of *virtual velocities*, the name given to it by John Bernoulli. Stevinus had employed the same principle towards the end of the sixteenth century in his investigations on the equilibrium of pulleys, and of bodies resting on inclined planes.

Lagrange derives his system of dynamics directly from d'Alembert's principle, and obtains a system of differential equations which are of a remarkably simple and symmetrical character, and at the same time extremely comprehensive. If the system consist of n material particles, then $3n$ co-ordinates will be required to completely specify the system, and $3n$ differential equations will be obtained, and these will be sufficient to determine the $3n$ co-ordinates at any given time in terms of the data of the problem, and therefore to determine the motion of the system completely, provided there are no connections, or constraints, between the particles. If such connections exist, then each of these will give rise to an equation between the co-ordinates. Suppose that there are m such equations. These will enable us to get rid of m out of the $3n$ previous equations, while taking account of the specified constraints. There will remain $3n-m$ of the original equations, which, together with the m equations of constraint, will still leave $3n$ equations, sufficient to determine the motion. The adaptability of Lagrange's method to a very wide range of physical problems is largely due to the extreme generality with which the co-ordinates may be chosen. In order to obtain any results from the mathematical investigation of a group of physical phenomena, we must be able to measure a

[1] *Mathematical Note.*—Lagrange's extension of Euler's principle was expressed in the form $\delta\sum m\int v ds=0$, signifying that the variation of the sum of all the integrals of the form $m\int v ds$ is equal to zero. It was pointed out by Jacobi that Euler's principle would not be subject to exceptions as mentioned in the text if expressed in the form $\delta\int v ds=0$. (See Appendix A.)

certain number of quantities in connection therewith. These quantities, whatever be their nature, may be taken as co-ordinates of the system. We shall, of course, endeavour to obtain, by observation, the values of a sufficient number of quantities to completely specify the system, as, unless this can be effected, a complete solution will be impossible. In some physical problems of the highest interest it may be impossible to determine a sufficient number of such co-ordinates to completely specify the system. But even in such cases the application of Lagrange's method may enable us to learn a great deal about the group of phenomena in question, for, if we can overcome all the mathematical difficulties, it will give us all the information that is obtainable from the data at our disposal. Let us, however, assume that we are dealing with a problem in which we have been able to determine the co-ordinates completely. Further observations will generally enable us to obtain a series of equations expressing relations between the co-ordinates due to mutual constraints. If all the actually existing relations of this kind are determined, in addition to a sufficient number of co-ordinates to specify the system, the resulting system of equations will completely determine the system for all time. Our theoretical knowledge of the group of phenomena will therefore be logically complete. The relations so obtained, however, will convey no definite ideas to a mathematically untrained mind, and even the trained mathematician will feel that he has obtained a much better grasp of the subject if he can translate relations of an unfamiliar character into comparatively familiar ones, such as the relations between moving masses or particles. This he will do by building up a system out of such familiar elements which will lead to similar sets of equations.

Any problem relating to moving particles or masses whose co-ordinates can be specified may be directly expressed in terms of Lagrange's system of equations; but in many classes of problems, such as those relating to molecular physics or electrical phenomena, the only possible method of expressing them in a dynamical form, except by the aid of unnecessary hypotheses, is in terms of the energy of the system. The principle of the "conservation of energy" asserts that, as the result of far-reaching experience, it is found that whenever energy disappears in one form it reappears in another without loss. This important principle was first stated in definite terms by George Green in the year 1837, and has been fully confirmed by all subsequent experience, so that it now ranks as one of the most firmly established facts of nature. It was, however, implicitly involved in the denial of the possibility of perpetual motion, which was the principle employed by Stevinus in

his investigations on equilibrium. This same principle of work was largely used by Huygens for the solution of numerous mechanical problems, and was implicitly contained in a scholium to Newton's third law of motion, which, as was pointed out by Lord Kelvin and Professor Tait, is practically equivalent to d'Alembert's principle. It also formed the essential basis of Euler's principle of "least action," and of the more complete and convenient expression of this principle which was formulated by Sir William Rowan Hamilton at the beginning of the nineteenth century. The energy of a dynamical system divides naturally into two portions, viz. the kinetic energy, which we will call T, and which, in the case of a system of particles, depends only on the masses and velocities of these particles; and the potential energy, which we will call W, and which depends only on the co-ordinates of the particles. The principle of the conservation of energy asserts that the sum of these is constant for any system, and is therefore expressed by the equation

$$T + W = \text{constant}.$$

If the system is one which it is convenient to express in terms of generalised co-ordinates, then W may be expressed in terms of the co-ordinates only, or, in mathematical language, will be a *function* of the co-ordinates only, while T will be a *function* of the co-ordinates and of their first derivatives with respect to the time. In the Hamiltonian form of the principle of "least action," the "action" which determines the motion is taken to be the difference between the kinetic and potential energies of the system. The principle then asserts that, when the system passes from its state at one instant of time to its state at any other instant, the path followed—in other words, the sequence of changes through which the system passes—will be such that the average value of this difference, during the interval of time occupied in the passage, will be smaller for the actual path than for any other possible closely adjacent path.

Whenever W and T are known functions of the co-ordinates this condition will enable us to obtain the Lagrangian equations which determine the motion. If W and T are not known functions, but if we can show that the former is a function of the co-ordinates only, and the latter a function only of the co-ordinates and their first derivatives with regard to the time, the possibility of a mechanical representation of the group of phenomena will be thereby demonstrated, but in that case there will be an infinite number of such representations which will fulfil the required conditions. Any system for which this principle holds will necessarily obey the law of the conservation of energy,

since this is implicitly involved in the condition that the potential energy of the system should be capable of expression in terms of the co-ordinates only.[1] The principle itself can be deduced as a consequence of d'Alembert's principle, which is, as we have seen, implicitly involved in the Newtonian system. Hertz, in his *Principles of Mechanics*, raised a practical objection to the use of Hamilton's principle in the explanation of natural phenomena. He points out that it involves the assumption that the natural system to which it is applied is of such a character that all its conceivable continuous motions are actually possible ones. This condition will necessarily be fulfilled if the potential energy is a function of the co-ordinates only; but, in problems of molecular dynamics, for example, the actual co-ordinates are practically infinite in number, and inaccessible to observation, and the co-ordinates employed are obtained from these by a process of averaging. This process implicitly involves the assumption that the system is not undergoing any change of constitution, and that its thermal conditions remain steady. In other words, the system is assumed to be a conservative or permanent one, viz. one which can be brought back to any previous condition by purely mechanical means, such as reversing the sensible velocities. It is often possible to separate out from a non-conservative system a portion which may be treated as forming a conservative one. Hamilton's principle may then be applied to this portion, while the forces arising from the actions of the remainder upon it must be separately estimated, and treated as applied from outside the conservative portion.

Lord Rayleigh, in 1873, made a very important extension of the principle of least action, which makes it possible in many cases to avoid the consideration of an external system of forces of this kind, which it may be very difficult to determine. He showed that, when the viscosity of a medium is of such a character that the frictional stress between any two of its particles is proportional to their relative velocity, in which case the geometrical similarity of the system is retained, then Hamilton's principle may be made applicable by the addition of what is called a dissipation function, which is expressible in terms of the co-ordinates and of the squares of the relative velocities. Even with this extension, however, Hamilton's principle cannot be regarded as rigidly applicable to any actual systems of natural phenomena.

[1] *Mathematical Note.*—Hamilton's principle is expressed in the equation

$$\delta\int_{t_0}^{t_1}(\mathrm{T}-\mathrm{W})dt=0\,,$$

where t_0 and t_1 are the initial and final values of the time.

It applies with rigid accuracy only to simplified presentations of them, but it certainly supplies us with very close approximations to the truth in most cases, if not always.

Hertz has propounded a still more general mechanical principle, which includes that of least action as a special case; but unfortunately he did not live to put it into application, and its increased generality makes its use more difficult. Even this more general principle, however, supposing that further mathematical developments should make its employment as convenient as that of the action principle, would not enable us to attack any actual natural problem in all its generality. We should still be obliged, though to a less extent than at present, to deal mathematically only with simplified presentations, or models, of natural phenomena, confining our attention to the limited groups of dynamical actions exhibited by these models. The gradual extension of such models by analytical methods, checked by constant comparison with the results of observation and experiment, supplies us with what we call mechanical theories of physical phenomena. Although ever growing, they can never become complete, and they must always remain liable to the possibility that some new observation, perhaps appearing at first to be of a trivial character, may necessitate a very extensive reconstruction. Some further remarks on the application of Hamilton's principle will be found in Appendix B.

Maxwell, in his treatise on *Electricity and Magnetism*, started from the conception of Faraday, that the lines of magnetic force tend to shorten themselves, and that they repel each other when placed side by side. He then expressed the value of the tension along the lines, and the pressure at right angles to them, in mathematical language, and proved that the state of stress thus assumed to exist in the medium would actually produce both the observed electrostatic forces and the observed forces on conductors carrying electric currents. He expressly points out (vol. i. p. 132) that he has not in any way accounted for this stress, or explained how it is maintained, and states that he has not been able to make the further step of accounting by mechanical considerations for these stresses in the dielectric. On the same page he observes, with reference to his demonstration of the possibility of explaining electrostatic phenomena in this manner:—

> "This step, however, seems to me an important one, as it explains, by the action of the consecutive parts of the medium, phenomena which were formerly supposed to be explicable only by direct action at a distance."

The investigation of the phenomena of electric currents is based on the capacity of an electric current for doing mechanical work,

independently of any external electromotive force maintaining the current. Capacity for doing work, he points out, is nothing less than energy, and the energy of an electric current is either of that form which consists in the actual motion of matter, known as kinetic energy, or of that which consists in the capacity for being set in motion, arising from forces acting between bodies placed in certain relative positions. This second kind of energy, which is called potential energy, is due to the action of what we call forces—that is to say, tendencies towards change of relative position. He observes that the electric current can be conceived as a kinetic phenomenon only. In the words of Faraday, "it is something progressive, and not a mere arrangement." Maxwell then shows that functions T and W can be found which represent the kinetic energy and the potential energy, respectively, of a system of conductors carrying electric currents, and which satisfy Hamilton's principle of least action. No assumptions are made as to the nature of the mechanism or of the connections of the parts of the system. The form and position of the conducting circuits are assumed to be determined by one set of co-ordinates, while a second set determines the motion of the electricity, and of anything whose motion is governed by that of the electricity. The existence of this moving matter, whatever may be its nature, is rendered necessary, on the assumption that there is no action at a distance, but only actions between contiguous portions of the medium, by the observed fact that currents in different circuits act upon one another.

Maxwell here deals with a limited set of phenomena, as set forth in Chapter I., and he shows that Hamilton's principle is applicable to them when expressed in terms of actions between adjacent portions of a continuous medium, and that a mechanical explanation of them is therefore possible on this basis. He himself only laid the foundations for such an explanation, but so solidly and so broadly that no demolitions, but only extensions and minor corrections, have yet been found requisite.

CHAPTER IV.

THE ETHER.

It is very frequently, but quite erroneously, assumed that the conception of an ether filling all space originated with Huygens, who was the first to propound seriously, and endeavour to develop, an undulatory theory of light. He was also the first to give a correct exposition of the general nature of the elasticity of a medium such as is required for the propagation of regular undulations, but the notion of the existence of such a medium has come down from a very early period of Greek physical speculations, and these were probably largely derived from very much earlier Aryan ones, which have only recently become accessible to modern European scholars. These early ideas can hardly have been founded on anything more definite than purely hypothetical speculation, but nevertheless it is interesting to note that they involved as a fundamental consequence the discrete or atomic structure of matter, and therefore the further consequence that matter is not divisible without limit. These early views of matter, as they were expounded by Democritus and Lucretius, bear a startlingly close resemblance, indeed, to those formed on the results of the latest scientific research. Both in the ancient and in the modern conceptions the ultimate reality is transferred from sensible matter to a uniform medium filling all space: all physical actions take place and are propagated in it, while the ultimate elements of matter consist of vortices or some other singularities in this all-pervading medium, moving through it with a continuity of existence, so that they can neither be created nor destroyed by any means conceivable to our minds. Lucretius regarded the ultimate material atoms as consisting of parts which were incapable of separate existence, and it is interesting to note this early foreshadowing of an essential part of the most recent atomic theory, as set forth in Chapter IX. of this volume.

The discrete atoms moving in empty space conceived by the physicists of the eighteenth, and the greater part of the nineteenth

century, presented the very serious philosophical difficulty of assuming a definite uniform limit to the divisibility of matter without affording any explanatory basis for the assumption. Lord Kelvin's beautiful theory of material atoms as consisting of vortex-rings of various shapes in the ether, which from this point of view would have to act as an ideally perfect, or frictionless, fluid, would fully account for the indivisibility of the ultimate atom, or sub-atom as it is now usually termed. The theory was based on the fact that, while fluids at rest do not possess any rigidity, portions of them may become rigid by being set in rapid motion, as is well illustrated by the smoke-rings which some smokers are so skilful in producing. They can also be visibly produced by very simple mechanical methods in air, water, and other fluids. The existence of rings produced in material fluids, which are the only ones of which we have cognisance, is only a brief one, as they are very soon dissipated, owing to the internal friction of the fluid. In the perfect fluid of mathematical physics, *i.e.* one in which there is no internal friction, von Helmholtz has shown that such a ring could not be called into existence by any method conceivable to us, and if existing, it could not be destroyed by any conceivable method.

Lord Kelvin's theory of vortex-atoms, however, although accounting for matter as a structure in the ether, and so tending to get rid of the duality of matter and ether, did not take any account of electric charges, and is therefore, in the light of our present knowledge, insufficient. The more general electron theory of matter set forth in Chapter IX. does not, however, exclude that of Lord Kelvin, as we shall there see.

It will be seen in later chapters that the development of electrical theory has for a long time been tending towards the recognition of an atomic distribution of electricity, and the electron theory, propounded by Lorenz, Voigt, Drude, Larmor, and others, is that the atomic charge is of the essence of each of the sub-atoms, of which an aggregation, moving round each other in unchanged orbits, forms the ordinary material atom. Two species of atomic charges, or "electrons," positive and negative, must be assumed, so that the most modern electrical theory is really bringing to life in a new form the once discredited hypothesis of distinct positive and negative electricities.

Now, a system of similar electrons arranged along the circumference of a circle, moving round it within the limits of speed required for stability, and constrained by the attraction of an electron of opposite sign at the centre, would constitute a vortex-ring in the surrounding ether. On the other hand, electron atoms of stable form may be imagined which would not constitute

vortex-rings, and there does not appear to be any reason for limiting the more general theory.

An atomic theory of matter, in which the atoms are regarded as discrete particles in an empty space, characterised by mere extension, cannot avoid the philosophical difficulty involved in picturing a definite uniform limit to the divisibility of matter. When, however, this geometrical abstraction is replaced by the concept of space as an entity forming the basis of the material universe, the difficulty disappears. Matter must then be regarded as consisting of differentiated portions of this entity, and the legitimacy of the ancient representation of such differentiations as permanent spins, which involves an atomic conception of matter, has now been demonstrated. We may, on the other hand, start by considering the atomic nature of matter as a fact deduced from observation and experiment, and then endeavour to picture to ourselves how these atoms can be linked together so as to be capable of the interactions which we observe between them. We then find ourselves driven to the assumption of a *plenum*, or medium completely occupying the intervening spaces, and transmitting the strains and motions determined by the atomic nuclei.

The first step towards utilising this concept of a continuous plenum in space in the correlation, or explanation, as it is commonly termed, of known phenomena, and in the prediction of phenomena as yet unknown, will consist in assigning a suitable system of properties to such a medium.

The most direct method of attacking this problem is to formulate in the first place a system of differential equations expressing the properties of such a medium in terms of as large a body as possible of known phenomena. The next step will be to test the validity and completeness of the system by comparing it with any phenomena which have not been taken account of in its construction, and introducing such extensions and modifications as may be necessary. If this process can be continued until all known phenomena in which the ether is concerned are included, the resulting system of equations will define the properties of the ether as fully as is possible in the existing state of our knowledge.

We may then endeavour to increase our knowledge by developing the system by mathematical analysis until new phenomena are indicated, and devising experimental methods of sufficient delicacy to enable us to determine whether the phenomena indicated by our system of equations are or are not actually existent. The discovery of such phenomena will afford further confirmation of the adequacy of our system of equations, while

the disproof of their existence, if this can be effected, will indicate a flaw, which must be remedied by making the necessary modifications in the system.

This course possesses the great advantage of guiding the work of the experimentalist into the most promising channels, and where it is possible it will certainly minimise the proportion of fruitless experimental research. It is, however, often arrested by mathematical difficulties which make it impossible, in the existing state of mathematical knowledge, to continue the development in its most general form. The mathematical physicist will then have to cast about for likely hypothetical assumptions, which may be followed up as long as they lead to results which are confirmed by experimental investigation, but must be abandoned when they come into conflict with it.

Most of the recent advances in explaining electrical and other physical phenomena from the standpoint of considering the ether as the ultimate reality, and as the basis of both electricity and matter, have been made in the manner indicated in the preceding paragraph. This is the analytical method. The earlier stages, however, were accomplished by a very much easier, although a longer, route. This easier route, by synthetic methods, can be rendered comprehensible to the general reader, and can be followed with facility by the student whose mathematical and physical knowledge is of quite modest dimensions, while the direct analytical route is only open to those who are familiar with the refined and difficult methods of modern mathematical analysis. The first step in the synthetic method will consist in assigning to the ether a set of properties which will enable it to account for the more obvious observed phenomena, and the next will be to examine the results of ascribing these properties to the ether, and compare them with other observed phenomena. If they are in accordance, we may then use the results of the ascribed properties to predict new phenomena, and test the validity of these predictions by experimental observations. When, at any stage in this procedure, discrepancies are observed, the ascribed properties will have to be so modified as to make their results agree with those obtained experimentally. As stated at the beginning of this chapter, Huygens was the first to set forth in a consistent form the general nature of the elasticity requisite to account for the transmission of regular undulations such as he assumed as the physical basis of the phenomena of light, and on this basis he succeeded in giving an explanation of the phenomena of reflection and refraction of light. The theory of the elasticity of the ether was greatly extended and developed by Fresnel, and although his scheme left various difficulties unaccounted for, it

proved of great value as a working hypothesis until it was replaced by a more complete and satisfactory one.

The most striking point about the ether is that it does not manifest itself in any way directly to our senses. It does not, like matter, oppose any sensible resistance to our movements, and it is clear at first sight that it must be capable of penetrating to the interior of any bodies capable of transmitting light even in the smallest degree. This would appear to suggest that it must either be, or approximate to, what mathematicians call a perfect fluid, viz. one in which there is no friction. Here, however, a difficulty arises. When sound, which is transmitted by ordinary matter, traverses a solid body, the vibrations take place in all possible directions. When, however, it is transmitted through a fluid, whether liquid or gas, the vibrations are entirely longitudinal, or, in other words, the particles taking part in the transmission have their movements confined to a backward and forward motion, parallel to the direction in which the sound is travelling. Fluids can only transmit vibrations of this kind, the reason being that they have no rigidity, or resistance to *shearing*, viz. to the sliding of one portion of the medium over another. Now, vibrations in directions transverse to the line of propagation are entirely due to this resistance to shearing, and therefore cannot be transmitted by fluids. The elasticity of a fluid is entirely what is called volume elasticity, or resistance against compression into a smaller volume. The vibrations in a light-wave are entirely transverse—that is to say, every particle of the medium remains throughout its motion always in the plane perpendicular to the direction of transmission of the ray of light. It appeared, therefore, to be necessary to assume that the ether, while behaving like a perfect, or almost perfect, fluid to bodies moving through it with comparative slowness, behaved like a very elastic solid with respect to the extremely rapid vibrations which constitute a ray of light; and this was the course taken by Fresnel. The non-existence of longitudinal vibrations in the transmission of light remained unexplained until Clerk-Maxwell showed that this was characteristic of electromagnetic waves, and suggested the probability of light-waves being simply electromagnetic waves having wave-lengths between the limits within which the human eye was capable of responding to them.

Newton was fully convinced of the necessity of an ether for the explanation of all physical actions, but he was prevented from accepting the wave theory of light propounded by Huygens, as he could not see his way, on this theory, to account for the sharp definition obtainable in the case of light-shadows. He endeavoured to overcome the difficulty in his corpuscular theory, which intro-

duced corpuscles moving in the direction of propagation, which corpuscles he assumed to be "adapted for reflection and refraction at equal intervals"[1] by means of etheric waves. Later on, Young and Fresnel showed that the rectilinear propagation of light, with its resulting sharply defined shadows, was a consequence of the extremely short wave-lengths of luminous vibrations. This discovery mâde the intervention of the corpuscles an altogether unnecessary complication, and there can hardly be a doubt that, if it had been made in Newton's lifetime, he would have abandoned it without waiting, as his successors did, for the demonstration of the inadequacy of the theory. Newton saw the necessity for an ether, but his followers, who adhered so long to the corpuscular theory of light, appear to have regarded it as making unnecessary the assumption of an ether filling the whole of the space occupied by the visible universe, and having properties of a character which had hitherto been quite unfamiliar. The corpuscular theory was only finally abandoned when it became possible to make comparative measurements of the velocity of light in media of varying density. It then became impossible to retain it further, as it demanded that the speed of transmission should increase with the density of the medium, whereas experiment showed a decrease, as required by the wave theory.

I have already mentioned that every body which is at all transparent to light—and all known substances are to some extent—must be permeated by the ether. Water and other fluids transmit light, but it cannot be the fluid which acts as the vibrating medium, for fluids have no rigidity, and are therefore, as already pointed out, incapable of transmitting the transverse vibrations of which light-waves entirely consist. Even in the case of transparent solids, the solid itself cannot be the medium of transmission; for the velocity of light, which is about 186,000 miles a second, is so great as to demand an elasticity in the medium very far in excess of that of any known solid, for it can be shown that the square of the speed of transmission is equal to the ratio of the rigidity of the medium to its density. Since the speed of transmission depends on the ratio of the rigidity to the density of the transmitting medium, considered as an elastic solid, it follows that the ether within different kinds of matter must be largely modified by the nature of the substance; for otherwise, the velocity of light-transmission through various kinds of matter would be independent of the nature of the matter, which, as mentioned above, is not the case. For example, the speed of a luminous wave

[1] It is difficult to form any conception of what Newton intended to convey by this very obscure remark, but it shows his recognition of the necessity of an ether for the explanation of optical phenomena.

passing through heavy glass is only about two-thirds of the speed in the free ether of space. The ether, considered as an elastic solid, must, therefore, either have its density increased, or its rigidity diminished, by the presence of the particles of glass. Fresnel, in his elastic solid theory of the ether, assumed that the density is increased by the presence of molecules of matter. Now, it is not sufficient for a theory of light to account for its transmission in the free ether and within the substance of a homogeneous mass of matter. It must also account for its passage, without any break in continuity, from one kind of matter to another, or between matter and free ether. In order to do this on any theory it will be found necessary that certain conditions should be satisfied at the surface separating the two kinds of matter from each other, or matter from free ether. If the ether is considered as an elastic solid, then the conditions will be that the displacement of the particles in any given direction must be equal on either side of the boundary, and that the elastic forces of restitution called into action by these displacements should also be equal on either side of the boundary. Each of these is determined by three conditions or co-ordinates, so that altogether six conditions have to be fulfilled. Now, two of these six conditions require for their determination the existence of the longitudinal waves, which, as previously pointed out, do not exist in the case of light. This is an indication of the fact that no elastic solid would transmit transverse vibrations without longitudinal ones, so that no elastic solid theory of the ether can be a complete one, as two of the requisite conditions are undetermined, so that, in order to determine them in any special case, it will be necessary to supplement the theory by special assumptions. Fresnel's theory does not even satisfy the four remaining conditions, for the assumption of a change of density introduces a discontinuity in the direction perpendicular to the boundary surface. This theory therefore satisfies only three of the six boundary conditions. Larmor suggests that it is also open to philosophical objection if the ether is to be regarded as the ultimate primordial medium completely filling the whole space occupied by the visible universe; for, if the density of the ether can be varied, it must be compressible. Its structure must therefore be molecular, and the molecules must be elastic; so that, if we are to adhere to our plan of assuming that every action between distant bodies is due to actual pushing or pulling of bodies actually in contact with them, a second ether will be required to explain the elasticity of the molecules of the first. This point will presently be further considered.

In the year 1837 a very interesting attempt was made by George Green to arrive at a general theory of the reflection and

refraction of light at the common surface of two non-crystallised media. It was not a successful attempt, owing to his having restricted the elasticity to that of an ordinary elastic body; but in spite of this, the paper possesses a very exceptional interest, both because it contains the first definite enunciation of the principle of the conservation of energy, and because it appears to represent the first definite realisation of the possibility which was inherent in Lagrange's dynamical system, of deducing all the phenomena of any mechanical system free from viscous forces from a single analytical function expressing the total energy of the system in terms of its configuration and motion. Green observes, in the introduction to the paper containing this attempt:—

"We are so perfectly ignorant of the mode of action of the elements of the luminiferous ether on each other, that it would seem a safer method to take some general physical principle as the basis of our reasoning, rather than assume certain modes of action, which, after all, may be widely different from the mechanism employed by nature . . . . and also lead to a much more simple process of calculation."

The principle is then enunciated in the following form:—

"In whatever way the elements of any material system may act upon each other, if all the internal forces exerted be multiplied by the elements of their respective directions, the total sum for any assigned portion of the mass will always be the exact differential of some function."

This function is simply what is now known as the potential energy of the system, and Green here simply makes a statement in mathematical terms, from which it follows as an immediate deduction that the sum of the kinetic and potential energies of the system is a constant quantity, which is the form in which the principle of the conservation of energy is usually expressed.

A couple of years after Green's attempt, James MacCullagh, who was not aware of Green's work, proposed to himself the problem of determining the form of this function for a continuous medium, so that it would lead to the laws discovered by Fresnel and others for the propagation, reflection, and refraction of light, including the observations of Brewster and Seebeck on the polarisation of light by reflection. He had previously attacked the problem by methods of pure geometry, and had arrived at results in accordance with the known facts of observation by assuming: (1) that the density of the ether is the same in all media; (2) that no loss of energy occurs in reflection and refraction; and (3) that the displacements in the incident and reflected waves, added together as directed quantities, or vectors (viz. in the same manner that velocities and forces are added), are geometrically

equivalent at the boundary to the displacements in the refracted waves, added in the same manner.

It will be convenient, before continuing our consideration of MacCullagh's contributions to a theory of the ether, to consider for a moment the nature of the vibrations in rays of ordinary and polarised light respectively. In general the vibrations constituting a ray of light take place in all possible directions in the plane of the wave. In order to fix the ideas, we will suppose for the moment that the vibrations consist of displacements of the particles of an elastic solid, as was assumed by Fresnel. Then, owing to the smallness of the excursion of each particle, it can be shown that the path of each vibrating particle will in general be practically an ellipse with its centre at the equilibrium position of the particle. These ellipses will, in general, vary in

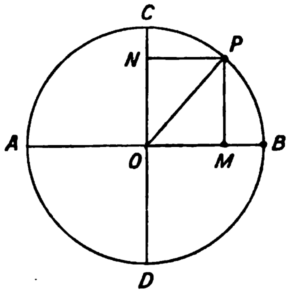


FIG. 1.

form, and in the directions of their longer axes. Let us consider a particle describing the simplest kind of ellipse, a circle, in the simplest possible way, viz. with uniform velocity. Let P (fig. 1) represent the position of this particle at any given moment. Through the centre, O, of the circle draw the two diameters AB and CD, at right angles to each other. From P let fall the perpendiculars PM and PN to the diameters AB and CD, intersecting them at the points M and N respectively. As P moves round the circle starting from A, and passing in succession through C, B, and D, and finally returns to A, it will be seen that M will move along the diameter AB from A to B and back again, while N will at the same time move along the diameter CD, from O to C, from C through O to D, and finally back to O. When P moves with uniform velocity in the circle, the motions of M and N are each said to be simple harmonic vibrations. It will be seen that these points do not move uniformly along the diameters, but that they are moving most rapidly when passing through the centre O,

and that their speeds gradually fall to zero at the extremities of their respective excursions, are then reversed, move with increasing speed until they reach the centre, and continue to move with diminishing speeds to the other extremities of their excursions. The motion of P is then said to be compounded of the motions of M and N, for, if a sheet of paper is made to move backwards and forwards with the simple harmonic motion of the point M, while the point of a pencil in contact with the paper moves up and down with the simple harmonic motion of the point N, then the path traced by the pencil upon the paper will be the circle ACBD. The lengths AB and CD, which in this case are equal, are said to be the amplitudes of vibration of the points M and N respectively. If the amplitude AB is increased, while CD is unchanged, and M and N are made to describe simple harmonic motions at right angles to each other as before, then the point P will describe an ellipse having AB as its longest, and CD as its shortest, axis. The time occupied by P in going through one complete cycle of changes from A to B and back again is called the period of the vibration, and the number of these cycles occurring in a second is called the periodicity or frequency of the vibration.

The angle AOP, measured in the direction of motion of the hands of a watch, is called the *phase* of the vibration executed by the point N; so we see that two simple harmonic motions of equal amplitudes, and differing in phase by 90°, compound into a uniform circular motion. If the difference of phase were 180°, or an odd multiple of 180°, the point P would remain at rest. Two superposed trains of luminous waves of equal amplitudes and maintaining such a difference in phase would therefore produce darkness, a phenomenon known as interference.

Now, it is possible, by allowing a beam of light to pass through certain crystals, or by reflection from crystals or metals, to partially or entirely break up the vibrations into simple harmonic vibrations at right angles to each other. When this is done completely, the light is said to be circularly polarised when the amplitudes of the mutually perpendicular vibrations are equal, and elliptically polarised when they are unequal. In the latter case all the particles taking part in the transmission describe similarly situated ellipses of equal dimensions. Absolutely homogeneous or monochromatic light is elliptically polarised.

It is further possible, by various simple methods, to remove one of these sets of vibrations from the beam, and allow the other only to pass. Such a beam is said to be plane polarised, as all its vibrations are then confined, not only to a plane at right angles to the direction of propagation, but also to a plane passing through this direction. Bartholinus observed, in the year 1670,

that when a crystal of Iceland spar was laid upon any object, and looked at with both eyes, two images of the object could always be seen, and that when the crystal was rotated about a vertical axis, the more refracted image revolved about the axis, while the other remained stationary. The phenomenon is known as double refraction, owing to the fact that a single incident ray gives rise to two refracted rays. Eight years later Huygens took up the investigation of this phenomenon, and found that while the less refrangible ray followed the ordinary law of refraction, that the sine of the angle of incidence bears a constant ratio (known as the index of refraction of the substance) to the sine of the angle of refraction, this was not true of the more refrangible ray, for which this ratio varied with the angle of incidence. He therefore called the latter the extraordinary ray, and the former the ordinary ray. There can be little doubt that, if Huygens had been aware that light-rays are transmitted by means of transverse, and not longitudinal, vibrations, he would have been led on to the discovery and explanation of the general phenomena of polarisation. His view of the ether as a perfect fluid led him to look only for the longitudinal compression waves which are alone transmitted by fluids. He succeeded, however, in explaining the phenomenon of double refraction as arising from varying elasticity of the ether in different directions in doubly refracting crystals, such as Iceland spar and quartz, assuming, in the first place, in order to account for the different properties of the ordinary and extraordinary rays, that the former were propagated by a wave with a spherical wave-front, while the wave-front of the latter was spheroidal[1] in form. He carried out a series of experimental observations of the most remarkable character, considering the very imperfect appliances at his disposal, the results of which were confirmatory of his assumptions. They were further confirmed by Wollaston in 1802, by the aid of greatly improved apparatus; and comparatively recent refined observations by Stokes, Mascart, and Glazebrook have fully established their correctness.

In the year 1810, Malus, who was engaged at the time in investigating the phenomena of double refraction, was casually observing, through a doubly refracting prism of quartz, a ray of sunlight reflected from a window, when he noticed that in certain positions of the prism the light disappeared. Malus gave the name of polarisation to this phenomenon, owing to rays of light possessing this property appearing to have different qualities in

[1] A spheroid is an ellipsoid with two of its three principal axes equal to each other—that is to say, it is an ellipsoid of revolution, or solid formed by the revolution of an ellipse about either its minor or its major axis. The former is called an oblate, and the latter a prolate, spheroid.

different directions, or, as we may say, in different *azimuths*, the variations representing polarisation taking place in planes perpendicular to the direction of propagation. Newton had observed many years before, with regard to doubly refracted rays, which are also polarised rays, that these rays appeared to have sides to them. Malus' observation directed the attention of Fresnel and Arago to the subject, and they very soon discovered that light could be polarised by reflection from a great number of very different substances for certain angles of incidence. They also found that, when a ray which had been polarised by reflection was transmitted through a crystal of Iceland spar, it acted as the ordinary ray when the axis of symmetry of the crystal was parallel to the plane containing the incident and reflected ray, or the plane of reflection. If, on the other hand, the axis of the crystal was perpendicular to the plane of reflection, the ray had the properties of the extraordinary ray.

Prior to these investigations Fresnel had assumed with Huygens that light-waves were of the compressional, longitudinal type transmitted by fluids, and it was the impossibility of accounting for the phenomena of polarisation which led him to try the assumption that the waves were transverse. He found that the results of this assumption were in accordance with the facts, and this led to the development of his elastic solid theory of the ether. An elastic solid was the only kind of body known which would transmit transverse vibrations, and though Fresnel did not succeed in accounting for the disappearance of the longitudinal waves, it was shown by Green, in his paper of 1837 referred to above, that these would be eliminated if the elastic solid medium were assumed to be practically incompressible. Fresnel and Arago found that, when light is polarised by successive reflection from a series of glass plates, the incident ray falling on each plate in succession at the polarising angle, there is very little loss of light in the second and subsequent reflections, if the successive reflections are in the same plane. If the second reflection is in a plane perpendicular to the first, the light still being incident on both plates at the polarising angle, then almost complete extinction is found to take place.

The plane of polarisation of a polarised ray is defined as a plane bearing the same relation to the ray as the plane of reflection would bear to it if the polarisation were due to reflection, so that, in the case of polarisation by reflection, the plane of polarisation is simply the plane of reflection. Sir David Brewster showed that, in singly refracting bodies, the polarising angle was such that the reflected and refracted rays were at right angles to each other, and that in this case the refracted ray was polarised in a

plane perpendicular to the plane of reflection, and therefore perpendicular to the plane of polarisation of the reflected ray. According to Fresnel's elastic solid theory, the displacements of the ether particles must take place in a plane at right angles to the plane of polarisation, while MacCullagh's theory leads to the conclusion that the displacements take place in the plane of polarisation. In his paper of 1839 MacCullagh directly attacked the problem of determining the expression for the potential energy on which the transverse rectilinear vibrations propagated through the ether must depend. He observed, in the first place, that vibrations of this character do not involve any compression of the medium. Therefore, in the case of a plane polarised wave in free ether, which is the simplest possible kind of light-wave, and therefore the most suitable for ascertaining the nature of the elasticity required, the particles of the medium will all be moving backwards and forwards in parallel directions. The strain at any point of the medium, due to the displacements of the neighbouring particles, will therefore be of the kind that would be produced by a rotation, or twist,[1] of the small element of volume surrounding the point, about an axis perpendicular to the direction of the displacements, and in the plane of the wave-front.

The amount of this twist will be proportional to the rate of change of the displacement in the direction of propagation. The elasticity of the medium must, therefore, be of such a kind that the deformations, or strains, due to this twist, will give rise to restoring forces, or stresses, in the medium, tending to produce a counter-twist which will bring the particles back to their positions of equilibrium after the passage of the disturbance producing the wave. This rotational, or twist, elasticity of the ether is of such fundamental importance in the explanation of electromagnetic phenomena that it seems advisable to illustrate the twist-producing effect of the displacements by means of a diagram.

Let the straight line AB (fig. 2) represent the line of propagation of a wave of plane polarised light in free ether, and let the plane of the paper represent the plane of polarisation. Every particle of the medium originally situated close to the line AB will then be executing simple harmonic vibrations in the plane of the paper, and at right angles to the line AB. In order to determine the simultaneous positions of the particles which, in the absence of any disturbance, lie along the line AB, we must describe a curve

[1] Any motion of turning about an axis is called a rotation, whether the turning be continuous or limited to a finite angle. I shall call the former kind of rotation *spin*, and the latter kind *twist*.

formed by compounding a simple harmonic motion in a direction at right angles to AB with a uniform motion along AB, representing the uniform velocity of light. This follows from the fact that all these particles are executing exactly similar vibrations; but, as the wave travels along AB in the direction of propagation, every successive particle starts from its position of equilibrium on AB when the wave reaches it, and therefore immediately after the one behind it. The simultaneous positions of the particles originally, when in equilibrium, lying along AB, will therefore be situated on the same curve as would be traced out by the motion of a single particle relative to the paper, supposing the latter to be moved in the direction BA, opposite to the direction of propagation, with the velocity of light. The curve ACODB is a curve of this kind, which, on account of its mathematical properties, is

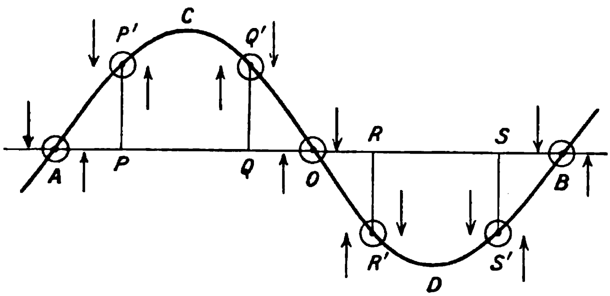


FIG. 2.

known as a sine curve. Sine curves of various shapes may be very simply obtained by cutting off obliquely, at any desired angle, one end of a wooden roller, wrapping a sheet of paper round the roller so that its edges just meet, and then cutting the paper with a sharp knife along the ellipse forming the edge of the cut end of the roller. If the paper is then spread out flat again, its cut edge will form a sine curve.

The points A, P, Q, O, R, S, B, originally in their positions of equilibrium, will, at the instant during the passage of the disturbance represented in the diagram, occupy the positions A, P′, Q′, O, R′, S′, B, respectively. The small circles drawn with these points as centres are to indicate the sections of small elements of volume of the medium which include these points.

Now let us consider the motion of the medium in the neighbourhood of one of these points, say the point P′. The portions of the medium close to P′, and on the right-hand side of it, will be moving upwards more rapidly than P′, and will therefore tend to drag

the volume element at P′ round in a counter-clockwise direction, as indicated by the upward-pointing arrow to the right. The portions of the medium close to P′, and on the left-hand side of it, will be moving upwards more slowly than P′, and will therefore again tend to drag the volume element round in a counter-clockwise direction, as indicated by the downward-pointing arrow to the left of P′. The directions of twist of the other selected volume elements are indicated by arrows in the same way. Since the twists are all in the plane of the paper, the axis of twist is perpendicular to this plane. The axis is therefore, as stated above, perpendicular to the direction of the displacements, and in the plane of the wave-front, for the plane of the wave-front is perpendicular to the direction of propagation.

MacCullagh then expressed the potential energy of the medium in terms of these elementary twists, assuming that the energy of the medium is entirely of a rotational character. He also assumed the density of the ether to be the same in all bodies, but its rigidity to be variable. He then obtained his equations of motion, which he found to give a consistent representation of all the known laws of propagation, reflection, and refraction of light. The expression of the energy as a quadratic function of the elementary twists, which was the form assumed by MacCullagh, was not the most general form possible, but, as it was found to lead to equations giving results in accordance with all the observed facts, it was not necessary to generalise it further. It follows from the equations obtained that, whether the medium be supposed to be compressible or not, no compression of the medium is involved in the motion.

Clerk-Maxwell observes, in his article "Ether" in the ninth edition of the *Encyclopædia Britannica*:—

"It is often asserted that the mere fact that a medium is elastic or compressible is a proof that the medium is not continuous, but is composed of separate parts having *void space between* them. But there is nothing *inconsistent with experience* is supposing elasticity or compressibility to be properties of every portion, however small, into which the medium can be conceived to be divided, in which case the medium would be strictly continuous."

The italics are mine. The great obstacle to the formation of any definite conception of the properties of a continuous substance is that all our sense-experience is of substance of a molecular nature. We may, as Maxwell observes, eliminate the necessity of void spaces *between* the smallest portions of a compressible ether, by assuming that these smallest portions are themselves compressible; but it appears to me that we are then driven to the conclusion that such void spaces must then exist *within* these

smallest portions, and this seems to involve the necessity of some kind of structure, either static or kinetic, in the ether itself, and that therefore we have not yet got down to the ultimate physical reality of continuous uniformity, or uniform continuity, whichever way we prefer to express it. If any physical phenomenon should be found to require compressibility in the ether, it might be accounted for by assuming the existence of certain kinds of ether vortices, or whirlpools. As mentioned before, there are no physical means of which we can form any conception which could start such vortices in a perfect frictionless fluid such as the ether must be, according to the electron theory of matter. These vortices would be distinct from matter, as they would not contain any electric charges, so that their aggregates would form a sort of intermediate stage between ether and ordinary matter. The complexity of observed phenomena appears to suggest some such further complexity in the ether. There do not, in the present state of knowledge, appear to be any indications of a necessity to make provision in our mechanical representations, or models, of the electromagnetic scheme of physical phenomena, for compressibility of the ether; but this does not enable us to dispense with the assumption of some kind of ether structure, either static or kinetic, more probably the latter, to form the necessary connecting links between ordinary matter and the ultimate *continuum* or *plenum*. The ether of MacCullagh's theory is essentially a simpler one than Fresnel's elastic solid-fluid, in that the only properties assigned to it are those which are requisite for the explanation of observed phenomena. It is, on the other hand, more difficult to conceive, because its properties bear no resemblance to those of any known substance. Sir George Stokes has pointed out that we can find a material analogy to the elastic solid-fluid ether in pitch, a substance which is highly elastic, and exhibits all the properties of a solid with respect to rapid motions; while with regard to slow motions it behaves as a fluid, for a slab of the hardest pitch will allow a leaden bullet laid upon its surface to sink through it, like any viscous fluid, that is to say, any fluid in which there is frictional resistance to flow. This illustration of the properties of the simple continuous plenum by comparison with a complex material substance such as pitch, cannot be considered as in any way explanatory, and is only referred to on account of its historical interest as the first crude attempt to form a conception of the properties of the ether. We shall see in Chapter IX. that the electron theory of the structure of matter satisfactorily accounts for its mobility through the ether. This absence of any material analogy for MacCullagh's ether appears to have been the main obstacle to

its general acceptation, and it seems to have been practically overlooked by physicists until G. F. Fitzgerald pointed out in 1880 that MacCullagh's equations were identical with those obtained by Clerk-Maxwell for the propagation of an electromagnetic disturbance, so that the properties of MacCullagh's ether were those required by the electromagnetic theory of light. If it had not been for the difficulty felt in forming a conception of such an ether without the assistance of a material analogy, it could hardly have failed to replace that of Fresnel, as it led to consistent representations of double refraction and other optical phenomena, which could only be interpreted, on Fresnel's theory, by means of assumptions which were mutually inconsistent. The only property assigned to MacCullagh's ether is resistance to twist of every small element of the medium, but this is inconceivable without the existence of some kind of structure, and without assigning to the medium, under the influence of luminous or other waves, some alteration in the distribution of twist throughout the area of disturbance. This distribution will be fully determined by the specification at each instant, and at each point of the medium, of the amount of this twist, and of the direction of the axis about which each elementary twist takes place. That is to say, everything is fully specified in terms of a single directed quantity, or vector, as such a quantity is called, which at any moment has a definite value at every point of the medium. Drude calls this quantity the *light-vector*. When MacCullagh's theory is applied to the representation of phenomena such as reflection and refraction, involving the passage of a ray of light between two material media, or between one material medium and the free ether, it is found that, when the density of the ether is assumed to remain constant, only four out of the six boundary conditions are independent, so that only four conditions have to be satisfied, and this is exactly the number required for the problems of reflection and refraction between crystalline media.

MacCullagh was fully impressed with the desirability of expressing the properties of the ether, as determined by his equations, in the form of what may be called a mechanical model, but he tells us that:—

"A mechanical account of the phenomena remained a desideratum which no efforts of mine had been able to supply."

Lord Kelvin, in 1889, showed the possibility of imagining a model, built up of ordinary matter, which is capable of representing, although in a very crude form, a medium having the property of elastically resisting twist in any direction, and Larmor has shown how this may be applied to show that MacCullagh's ether

is a legitimate dynamical concept. Before describing this, however, I must warn my readers against supposing that such an illustration can be anything more than an assistance in forming a mental picture of the scheme of mathematical relations involved. It is not intended to suggest that the resistance of the ether to twist may be due to the presence in it of such structures. The view that I am putting forward, and which appears a very probable one, is that the molecules of matter are structures in the ether, and formed from the ether. The ether, whatever it may be, is most certainly not a structure built up of matter.

Suppose the medium to be built up of an aggregation of frictionless spheres each studded over symmetrically with a small number (*e.g.* four) of short frictionless spikes. This will form a crude, but conceivable, representation of a perfect fluid. These spikes will then maintain continuity in the motion of the medium, without the aid of viscosity, and they will also compel each sphere to participate in the twist of the portion of the medium containing it. In order to obtain the property of elastic resistance to twist, Lord Kelvin imagined that each sphere contained three gyroscopic tops, consisting as usual of a heavy disc spinning about an axis through its centre and perpendicular to its plane, this axis being pivoted within a ring in such a way as to form a diameter of the ring. If three such rings are pivoted within each sphere in such a way that the axes about which they are capable of rotating are in three directions at right angles to each other, while the axis of each ring is at right angles to the axis about which the disc rotates, this result will be attained, and the resistance to twist will be independent of its direction. The elasticity will be what is termed isotropic. This result will only hold good provided the total displacements of the axes of the gyrostatic tops are small, but this is all that is required to confer upon the medium the necessary rotational elasticity to enable it to transmit vibrations of small amplitude, corresponding to the luminous and other electromagnetic vibrations transmitted by the ether. We shall see in Chapter IX. that there will be no twist in the ether except in the presence of electric force. We shall also find it necessary, on account of the high speed of propagation of light and other electromagnetic waves, to assume a high value for the inertia of the medium, and therefore an extremely high elasticity, so that the strongest electric forces will only give rise to very slight twists in the ether. The observed high value of the inertia of the ether of space, which was referred to in Chapter II., shows that this assumption is legitimate. These conditions may be represented in our model by giving a

high moment of momentum[1] to the gyrostatic tops, by employing heavy discs spinning with a very high velocity. The model can therefore be made to represent a field of steady electric force lasting, without sensible decay, for any assigned length of time, although not for ever.

According to the view of the nature of matter here suggested, the ether itself must not merely approximate to, but actually be, an ideally perfect frictionless fluid; for if there were any trace of friction in it, the whole physical universe would, if our theory of the constitution of matter is the true one, ultimately dissolve into the ether without leaving a trace behind.

[1] The moment of momentum of a mass spinning about an axis is obtained by multiplying the momentum of each of its particles by the distance of the particle from the axis, and adding the products. In the case of a continuous mass this is, of course, effected by integrating the product throughout the mass.

CHAPTER V.

THE ETHER AS A FRAMEWORK FOR ABSOLUTE MOTIONS.

LORD KELVIN's illustration of gyroscopic rotational elasticity raises what may seem a philosophical difficulty, in that the elementary portions of such a medium as is represented in his model appear to offer a resistance to absolute, and not merely to relative, twists, and apparently, therefore, indicate absolute directions in space. This is not by any means the first appearance of such a difficulty in connection with motions of rotation, but in the present instance it presents itself in an exceptionally fundamental form. Consider, for example, such problems as the shape assumed by the surface of a liquid in a rotating bucket; the figure of the earth; the resistance of a spinning body such as a top, and of the earth itself, to any displacement of its axis of spin; the resistance of a swinging pendulum to the displacement of the plane of swing. A freely suspended pendulum of considerable length and mass was employed by Foucault as a means of demonstrating the actual, or absolute, rotation of the earth, its movements being such that they could be explained most simply, and, on the basis of the accepted principles of mechanics, could be explained only, on the assumption that the pendulum tended to conserve the absolute aspect in space of its plane of swing. Newton regarded phenomena of this kind as definitely indicating absolute rotations, and therefore absolute directions in space, and he has expressed his opinion to that effect in the most unequivocal terms. Amongst the ablest critics of Newton's position on this subject I may mention E. Mach and H. Poincaré. Both these physicists have discussed the question in considerable detail, the former in his treatise on *The Science of Mechanics*, and the latter in *La Science et l'Hypothèse*; and both object to what they conceive to be Newton's position, that he attempts to explain relations between phenomena by means of the notions of absolute direction and absolute rotation, involving the notion of absolute space, which is physically inconceivable to the relatively

constituted human mind. This objection would appear to be a well-founded one, if the meaning attached, in Newton's mind, to the expression absolute space were identical with that attached to it by his critics, and which they have defined with perfect clearness. Space for them is the geometrical abstraction of mere extension, and in this sense absolute space is an expression which does not appear capable of representation in the form of a definite mental concept. The difficulty appears to me to be of very much the same order as the mental representation of direct action at a distance, but of a much lower order of difficulty than that involved in the question of the relations between mind and matter. Even if the objection be well founded, it cannot be accepted as warranting the conclusion that there cannot be physical phenomena indicative of absolute motion and absolute direction in space, for this would involve the assumption that our minds are to be taken as measures of the possibilities of the physical universe. On the other hand, the replacement, if possible, of such vague and difficult ideas by the simpler ones of relative motion and relative direction in space is in every way preferable, and constitutes a distinct advance in physical theory. Other writings of Newton's, however, demonstrate, I think, that this was not his conception of space, but that he regarded it as occupied by a physical reality, the ether, and that this physical reality formed the requisite framework of reference, though perhaps but dimly outlined, in his mental consciousness. It may even have been too dimly outlined to make it possible for him to crystallise the concept into a statement which would have carried conviction to the minds of his contemporaries.

Poincaré instances the case of a planet, such as the earth, so veiled in clouds that no external bodies are perceptible, and argues that in such a case we could not infer with certainty that the planet was rotating, but only that such an assumption would provide the simplest basis possible for a system of mechanics. In his later work, *La Valeur de la Science*, he qualifies his contention by the admission that the assumption, in addition to its simplification of the system of mechanics, would render it possible to establish so much wider a correlation amongst physical phenomena, that it might be considered as of the same order of probability as the reality of the external world. In other words, his conclusion is that practical, though not theoretical, certainty of the rotation would be attainable by the inhabitants of such a world. I am unable to distinguish this from an abandonment of his position, for no thinking human being would endeavour to maintain so untenable a proposition as the possibility of attaining theoretical certainty in any conclusions derived from experience.

In the same volume Poincaré gives utterance to an almost pathetic lamentation at the trend of recent additions to our physical experience towards the shifting of the fundamental principles of physics from the well-explored and well-tested foundation of ponderable matter to the comparatively unknown and untried foundation of the ether. Although he does not make any reference to this in his qualification of his original contentions with regard to motions of rotation, I cannot avoid the conclusion that the qualification is due, whether consciously or unconsciously, to his consideration of this transference and its consequences.

Mach's argument is both interesting and instructive. He points out that, when we make the statement that "a body preserves unchanged its direction and velocity *in space*, our assertion is nothing more or less than an abbreviated reference to *the entire universe*." In another place he observes that:—

"When we reflect that the time-factor that enters into the acceleration is nothing more than a quantity that is the measure of the distances, or the angles of rotation, of the bodies of the universe, we see that even in the simplest case, in which apparently we deal with the mutual action of only *two* masses, the neglecting of the rest of the world is impossible. Nature does not begin with elements, as we are obliged to begin with them. It is certainly fortunate for us, that we can, from time to time, turn aside our eyes from the overpowering unity of the All, and allow them to rest on individual details. But we should not omit ultimately to complete and correct our views by a thorough consideration of the things which for the time being we left out of account."

I must give one more extract from Mach, owing to its direct bearing upon our present point of view. He points out the possibility that, in the consideration of the motion of a selected body, other isolated bodies might be found to play only a collateral rôle in the determination of the motion of the selected body, and that this motion might be determined by a *medium* in which this body exists.

"In such a case we should have to substitute this medium for Newton's absolute space. *Newton certainly did not entertain this idea.* Moreover, it is easily demonstrable that the atmosphere is not this motion-determinative medium."

The italics in the last sentence are mine. I consider this statement of Mach's to be a mistaken one, arising most probably from his having confined his attention to the *Principia*, which only contains a brief and casual reference to the necessity of an ether. Newton's conviction of the necessity of an ether for the explanation of gravitational phenomena is definitely stated in his letters to Bentley, and his recorded statements with respect to the optical necessity of such a medium were collected by Thomas

Young, and printed in his paper on "The Theory of Light and Colours," which was communicated to the Royal Society in 1801. Mach continues:—

"We should, therefore, have to picture to ourselves some other medium, filling, say, all space, with respect to the constitution of which and its kinetic relations to the bodies placed in it we have at present no adequate knowledge. In itself such a state of things would not belong to the impossibilities. It is known, from recent hydrodynamical investigations, that a rigid body experiences resistance in a frictionless fluid only when its velocity *changes*. True, this result is derived theoretically from the notion of inertia; but it might, conversely, also be regarded as the primitive fact from which we have to start. Although practically, and at present, nothing is to be accomplished with this conception, we might still hope to learn more in the future concerning this hypothetical medium; and from the point of view of science it would be in every respect a more valuable acquisition than the forlorn idea of absolute space."

The legitimacy of Newton's contention as it stands is very ably supported by B. Russell, in his *Principles of Mathematics*, a work which I have only seen since writing the earlier portion of this chapter, and I am glad to have the opportunity of outlining his argument, which is most instructive and interesting. It will also assist in further emphasising the explanation given in Chapter III., that in the application of mathematics to physical phenomena it is only possible to deal directly with ideal models or representations, and that the validity of the results obtained can only be ascertained by their comparison with data derived from actual observation.

The machinery of pure mathematics itself is concerned only with formal relations, and is essentially a development of pure logic. The similarity of the relations between the series of points on a line and the series of instants in an interval of time enables similar mathematical processes to be employed in dealing with spatial relations as with relations in time, as was pointed out in Chapter III.; and pure mathematics concerns itself only with the relations, and not with the question as to whether the units dealt with are points or instants.

Russell's work is devoted to showing that pure mathematics rigidly interpreted in this sense, and including arithmetic and analysis, the various systems of geometry, and what are called rational dynamics, can all be strictly deduced from a very small number of fundamental logical principles. The theory of arithmetic is first arrived at in this manner, and it is shown that the recent development of arithmetical theory enables all the propositions of the other branches of pure mathematics to be derived from arithmetical considerations alone. It must be

borne in mind that pure mathematics in this strict sense does not in any way concern itself with the reality or otherwise of the axioms employed. This is a question for applied mathematics.

For example, the purely mathematical development of Euclidean geometry does not concern itself with the question whether Euclid's axioms apply to our actual space, its province being rigidly confined to deducing the consequences of the axioms assumed.

In this sense, it is shown that strictly demonstrable mathematical concepts of absolute time, absolute position and direction in space, and absolute motions of translation and rotation are attainable, and it is on this ground that Russell upholds Newton's contention in its actual verbal expression, and without any such suggestion as I have made, that there may have been an implied though unexpressed reference to the ether as a framework. Now, I may say at once that I find myself in complete agreement with Russell, and fully admit the validity of his contention with regard to the various mathematical spaces of purely geometrical systems and the absolute motions of rational dynamics. Newton, however, was not dealing with the purely mathematical concepts of rational dynamics in his references to absolute motion, but with the physical concepts of applied dynamics—a very different matter.

I have no difficulty in forming a mathematical concept, for example, of a space of 4, or 5, or even of n dimensions, since all such spaces are defined by means of perfectly definite mathematical relations; but I should be utterly incapable of forming anything that could possibly be called a physical conception of a 4-dimensional space, even by the aid of such items of knowledge as that a being capable of 4-dimensional motion could enter a closed room in our 3-dimensional world without passing through its boundaries.

The infinity of our own Euclidean space, and the fact that figures can be moved about in it without undergoing any changes in size or shape, make it extremely difficult to form anything approaching a definite physical concept of absolute positions and directions which there are no means of marking, as all parts of space are exactly similar. The most effective way I can imagine of partially surmounting this difficulty is to replace Euclid's axiom with respect to parallel straight lines by one which makes them meet when produced in either direction at a great but finite distance. If the distance were sufficiently great, we should not be able to distinguish between such a space and the Euclidean one; but it would transform it into what is known as an elliptic space, which would be of finite extent, and, moreover, all figures moved about in it would change in size and shape as they moved from one place to another. If the distance selected were sufficiently great, these

changes would be imperceptible to us, and in fact we cannot say that this is not the real nature of our space.[1] The character of these changes, in a much exaggerated form, would resemble those observed in the reflections of moving objects in the interior of an ordinary metallic dish-cover. The conception of such changes, even though so small as to be imperceptible, will greatly assist us to fix our thoughts on a selected point in space, since space is now, not only finite, but differs in character from point to point.

When we think of the space as filled with something, such as the ether, it seems to be much easier to think of position or direction relatively to it, even if we think of the ether only as a perfectly uniform continuous medium; and it becomes easier still when we think of space as full of ether whirls or spins which have to be traversed in moving from point to point.

[1] Indications of the finite extent of our universe of matter and ether, which are considered in Chapter XXIII., appear to suggest the objective reality of this concept.

CHAPTER VI.

RELATIONS BETWEEN ETHER AND MOVING MATTER.

We saw in Chapter IV. that the explanation of optical phenomena requires that all material bodies shall be freely penetrable by the ether. We know also, as a matter of ordinary observation, that, whatever relations may exist between ether and matter, they are not of a character to produce any perceptible direct effect upon our sensory organs. If any resistance were offered by the ether to the motion of material masses through it, the most immediately obvious means of detecting it would be found in its effects in retarding the motions of the heavenly bodies. The motions of the planets and their satellites can be computed so accurately that extremely small deviations due to the presence of a resisting medium would, in time, be brought within the observational powers of astronomical measuring instruments. Such effects have been looked for, and it was at one time thought that indications of the presence of a resisting medium were afforded by the motions of some comets, notably Encke's comet. At present, if the existence of such a retardation were established, we should ascribe it to the presence in space of meteorites and cosmic dust (see Chapter XIX.), or to gravitational effects considered in Chapter XXIII., and not to the presence of the ether. If any such resistance were offered by the ether to the motions of the heavenly bodies, it would necessitate a certain dragging along with it of the ether by a moving body, and such displacements would lead to optical effects which afford much more sensitive tests than the change of motion due to resistance.

The first of these which we shall consider is the astronomical phenomenon of the *aberration of light*, which was discovered by Bradley in the year 1727. A small proportion of the so-called fixed stars are near enough to the earth to appear sensibly displaced on the celestial sphere when viewed from opposite extremities of a diameter of the earth's orbit, a distance of about 186,000,000 miles, viz. when the terrestrial observations are made at intervals of six months. The angle between the two directions in which the star is seen—that is to say, the angle

subtended at the star by the diameter of the earth's orbit—is called the annual parallax of the star. It is easy to see that the apparent displacement due to the annual parallax will take place along a great circle[1] passing through the apparent positions of the sun and the star in the celestial sphere. Bradley was endeavouring to detect traces of annual parallax in several stars which passed near his zenith,[2] when they would be in the most favourable position for its measurement. His observations showed displacements in the position of each of the stars under observation, and these displacements had the expected complete period of one year. The displacements, however, were not directed towards, or away from, the sun, as they would have been if they were due to annual parallax. They were in a direction at right angles to this—that is, they were in a direction parallel to the earth's motion in its orbit. After many attempts to account for these displacements, the explanation is said to have been suggested to Bradley when looking at a flag floating from the masthead of a ship. When the ship changed its course he noticed that the flag flew in a different direction. A more direct illustration is afforded by perpendicularly falling raindrops, which appear to drive towards the face of an observer moving forwards relatively to the air which carries them. This explanation would hold good on any theory of light propagated in rays travelling with a finite speed. On Newton's corpuscular theory it would require that the luminous particles should not be sensibly affected by the earth's gravitation; and on the wave theory it would require, either that the ether should not be sensibly disturbed by the earth's motion through it, or that there should be some special adjustment leading to the same result.

Arago made some very careful measurements of the amount of refraction of a ray of light from a star, traversing a glass prism, when the direction of propagation was the same as that of the earth's orbital motion, and when it was in the opposite direction, respectively. He was led to make these experiments owing to the fact that the corpuscular theory of light indicated a difference in the speed of transmission in glass and in air, and therefore he argued that the aberration of its path ought to exhibit a corresponding difference. His calculations indicated that a difference of about a minute in the angle of refraction was to be anticipated, an amount which his apparatus was perfectly capable of detecting. The result of the experiments was that no difference could be detected. From this he made the perfectly valid inference of the

[1] A great circle on a sphere is a section of the spherical surface by a plane through its centre.

[2] The observer's zenith is the point of the celestial sphere which is vertically above him.

inadequacy of the generally accepted corpuscular explanation of the phenomenon of aberration. He therefore inquired of Fresnel whether he could account more satisfactorily by means of the wave-theory for the non-effect of the earth's motion on the refraction of light. Fresnel, in his reply, first pointed out that the elucidation of this point was the more important in that similar considerations would apply to light from terrestrial sources, since the velocity of the waves is independent of the motion of the body from which they emanate.

It will be of interest to point out at this stage that, although the truth of this statement of Fresnel's as to the speed of propagation being independent of the motion of the source of light is confirmed by our present knowledge, yet Arago's experiment would have shown a difference if he had been able to employ measuring appliances of the sensitiveness now available. This would not have been due to any alteration in the aberration, or in the velocity of propagation of the light, but to an alteration in the wave-length, due to the different relative motions of the star and the observer in the two cases. When the prism was being carried towards the star by the earth's orbital motion, more waves per second would fall on the prism than when the latter was being carried in the opposite direction. Now, as we shall see in Chapter XII., these impulses would give rise to a new series of ether waves within the prism, the wave-lengths of which, being determined by the number of impulses falling on the prism per second, would therefore be shorter, and consequently the refraction greater, in the former case than in the latter. This is known, from the name of its discoverer, as the Doppler effect, and has been applied by Sir William Huggins to the determination of the speed with which stars are approaching, or receding from, the earth. A simple analogue is afforded by the case of a regiment of soldiers marching in ranks with regular intervals and at a uniform speed past an observer. If the observer move in the direction opposite that in which the regiment is marching, he will pass more ranks in a given time than if he remain at rest; while if he walk in the same direction with a slower speed than the regiment, he will pass fewer ranks in a given time than if he had remained at rest.[1]

A far more complete analogue is found in the familiar observation that the note emitted by the whistle of a locomotive appears to be raised when approaching, and flattened when receding from, an observer. In the former case the drum of the ear receives fewer, and in the latter case more, impulses per second than when the train is at rest. It is the number of impulses falling on the

[1] See also Poynting and Thomson's *Heat*, p. 338.

drum in a second which determines the motion of the air within the ear, and, therefore, the pitch of the note.

Fresnel then goes on to observe that it would be easy to explain the fact that the reflection and refraction of light are unaffected by the motions of the observing apparatus relatively to the source, if the ether surrounding the apparatus could be considered as carried along by the earth in its motion, but that the facts of aberration appeared to negative the possibility of this being the case. In view of these facts, he was driven to the conclusion that the earth and other bodies must be freely permeable by the ether, and capable of moving through it without giving rise to displacements of perceptible amount. Assuming, therefore, the velocity of ether waves outside a transparent body to have its normal value, the problem to be solved was to determine the possibility of such a modification of the velocity within a medium moving through the ether as would make the laws of refraction (and reflection) identical with those for a body at rest. Fresnel easily showed that this would be the case, in Arago's experiment, if the velocity within the glass were increased by the amount $v\left(1-\frac{1}{n^2}\right)$, where v is the velocity of the glass in the direction of propagation of the light, and n is the index of refraction of the glass. The modification in the case of the reflected rays would necessarily follow from the modification of the refracted rays. Arago's experiment was repeated with much more delicate apparatus, but with the same result, by Clerk-Maxwell, and in 1867 Maxwell worked out the theory of Fresnel's law more completely. Taking V as the velocity of light in air, V′ as its velocity in the denser medium, v' as the amount by which V is altered by the motion of the ether relative to the medium, and v the velocity of the ether in air relative to the observer, it is found that, neglecting squares and higher powers of v'/V, that is to say, to the first order of small quantities, the absolute velocity of the ray would have to be increased by $v(1-V'^2/V^2)$. In an isotropic medium (viz. one which has identical properties in all directions) V'/V is the index of refraction, so that the law for such media is as given by Fresnel. The latter attempted, but not very successfully, to account for the law on a dynamical basis. We shall see in Chapter IX. that such a basis for the law is provided by the electron theory of matter.

Boscovitch had, at a much earlier date, suggested that, since the velocity of light in water differed from that in air, a difference in the aberration of a star might be detected if the star were first observed in the ordinary way through an astronomical telescope, and then through a telescope filled with water. This experiment

does not appear to have been carried out until, towards the end of last century, Sir George Airey had such an instrument temporarily erected at Greenwich Observatory. No difference, however, was observed, showing that the different aberration of the ray in water was exactly compensated by the Fresnel modification of the refraction on the passage of the rays of light into that moving medium.

Since Arago's experiment, numerous investigations have been made, in many cases with very delicate apparatus, to decide whether other optical phenomena were affected in any manner by the earth's motion. The results have all been decidedly negative, except in the case of experiments made by Fizeau, in which he thought he could detect some evidence of displacement, due to the earth's motion, of the plane of polarisation of light passing through a pile of glass plates. Fizeau himself, however, regarded this result as extremely doubtful, owing to the difficulty experienced in eliminating disturbing effects. In 1893 Sir Oliver Lodge attempted to detect some motion of the ether between two steel discs close together and set in rapid rotation in opposite directions. These observations, which were of a very refined character, gave decisively negative results.

A brilliant attempt was made by Sir George Stokes, more than half a century ago, to develop the theory of an ether which should be carried along with the earth in its motion, as he was not prepared to admit the validity of Fresnel's view that the ether could pass freely through the interspaces of material bodies. This was a very formidable difficulty, having regard to the views then held as to the constitution of solid masses of matter, and was only adopted by Fresnel as a refuge from the still greater difficulties which appeared to be introduced by the assumption that the ether was carried along by the earth in its motion. The difficulty of imagining the possibility of the ether passing freely through solid bodies, as the wind passes through a grove of trees—to use a simile employed by Thomas Young—is entirely removed, in common with many other difficulties, by the electron theory. This we shall see later. It is worth while, however, to enable the reader to see for himself that even Stokes' ability was unable to surmount the difficulties inherent in theories of a moving ether. The method of approaching the problem was suggested to Stokes by hydrodynamical investigations upon which he had been engaged, and his starting-point was a theorem of Lagrange in which he showed that no differential spin, that is to say, spin of portions of the medium relatively to other portions, can ever arise in an ideally frictionless fluid which was originally at rest. This absence of differential spin was essential to enable the

phenomena of aberration to be accounted for; for in its absence there would be no slewing round of the wave-front of a train of luminous waves in its passage from ether at rest to ether in motion, and therefore the propagation of the rays would continue to be rectilinear, just as if the ether were at rest. The waves entering the observer's telescope would therefore be affected relatively to him with the full aberrational change of direction due to his motion, exactly as in the case of a moving particle. Stokes assumed the medium to have the properties of a frictionless fluid for the comparatively slow motions of bodies through it, and the properties of an elastic solid for the extremely rapid vibrations involved in light-waves. The possibility of these properties coexisting in the same medium was illustrated by the behaviour of pitch, as was mentioned in Chapter IV. Here a difficulty arose in the fact that spinless motion, if once established in a frictionless fluid, would become unstable, and the fluid would not maintain its continuity. Stokes attempted to overcome this difficulty by the assumption that the fluid had infinitesimal traces of viscosity, as he had recently succeeded in establishing mathematically that the steady motion of a sphere through a viscous fluid maintains its character however small the viscosity. Moreover, spinless flow gives rise to no viscous reactions within the mass of a viscous fluid. This would enable ideally frictionless solids to traverse the medium without giving rise to spin, in spite of the slip at their surfaces. The ordinary frictional surfaces of bodies would, however, give rise to successive layers of spin, spreading outwards from the surfaces in a manner similar to the diffusion of heat by conduction from a hot body. This difficulty he overcame by showing that the elastic solid qualities of the medium required for the transmission of the transverse luminous vibrations would maintain the spinless nature of the flow, because the spin would now be propagated away from the moving surfaces by transverse waves of displacement, instead of by diffusion, as in an ordinary fluid, and the speed of light required so high an elasticity that any incipient spin would be immediately shed off from the surface, and dispersed.

This theory, beautiful as it is, does not, as Larmor points out, provide a medium of the nature required for the representation, nor does it obviate difficulties which appear to be inherent in the assumption of a moving ether. The transparency of the ether must be sufficient to convey the light of the most distant stars, which would appear to require absolute mathematical transparency in place of the approximate transparency of such a medium as that of Stokes. Moreover, this imperfect transparency, due to the assumption of traces of viscosity, would necessarily involve

the decay of all structures existing in the ether, so that such an ether could not form a basis for matter. Even if this difficulty were overcome, another equally serious one would confront us. The phenomena of aberration and the negative results of all optical attempts to obtain evidence of motion through the ether would be fully accounted for if it could be assumed that the earth's motion gives rise to differentially spinless flow in the surrounding ether, but of such a character that in all regions extending from the earth's surface to the greatest distances at which exploration is possible, the ether is carried along bodily with the earth. We could not make this assumption without also assuming the ether to be carried along by other heavenly bodies in the same way, and then, whether we assumed the intervening ether at a distance from all such bodies to be at rest or not, it would be impossible, without discontinuity, to connect these states of motion by any possible spinless motions of the intervening ether. Similar difficulties would be encountered in the case of moving bodies on the earth's surface, such as Lodge's discs, for example. Larmor observes that we might be tempted to replace the absolutely spinless motion of the surrounding ether, which involves slip at the surfaces of solids traversing it, by motion with very slight spin such as would obviate the slip. This, however, would not be consistent with the phenomena of aberration, for, the smaller the spin, the greater the distance to which it must extend, while the resulting aberration is proportional to the product of these quantities.

We shall see in Chapter IX. that, quite independently of any special assumption as to the relations between ether and matter, such as that of the electron theory, the general principles of the Faraday-Maxwell representation of the transmission of electromagnetic actions by means of a medium, as opposed to the theory of action at a distance, show that any visible optical effects due to the relative motions of ether and matter must, if they exist at all, be exceedingly small. We shall see that no effects proportional to the ratio v/V, viz. of the velocity of translation of the matter relatively to the ether, to the velocity of light in free ether, can be anticipated, but only effects proportional to its square, or higher powers. The highest value that can be attained by v will be that due to the motion of the earth through a quiescent or stagnant ether. The principal component of this, amounting to about 30 kilometres a second, will be the component due to the earth's orbital motion. The component due to the earth's rotation is not quite half a kilometre a second, and may be neglected in comparison with this. There will also be a component of unknown amount due to the motion, discovered by Herschel, of

the whole solar system through space in the direction of the constellation Hercules. This will increase or diminish that due to the orbital motion, according to the position of the earth in its orbit at the time of observation, causing an increase when the orbital motion has a component in the direction of Hercules, and a decrease when there is a component in the reverse direction. The experiments made to look for an effect depending on the existence of a velocity v of the earth relative to the ether have extended over long periods, and many of them, especially the celebrated Michelson-Morley experiment, which will be considered presently, besides extending over long periods, have been carried out with a full knowledge, on the part of the observers, of the possibility of the motion of the solar system through space causing a diminution of the effect at certain seasons of the year. We may therefore be certain that during at least half the year the velocity v is not less than 30 kilometres a second, if the ether is stagnant, and we have seen that the assumption that it is carried along by the earth would involve us in insuperable difficulties. Now, the value of V is about 300,000 kilometres a second, so that the ratio v/V will not be less than one ten-thousandth. Since any effect to be looked for cannot be proportional to v/V, but at the most only to v^2/V^2, that is, to one-hundred millionth, it will be seen that only an extraordinarily delicate method of observation could be expected to lead to any perceptible result. It was suggested by Maxwell that it might be possible to detect a difference in the time of propagation of a ray of light between two points at a fixed distance from each other when the straight line joining them was placed, first in the direction of the earth's orbital motion, and then in a direction at right angles to it. The experiment was made by Michelson in 1881, and no effect whatever was observed. The result showed conclusively that there could be no effect of the first order as regards the small quantity v/V; but the method was not considered sufficiently sensitive to prove beyond all doubt that there was no second-order effect, viz. one proportional to the square of v/V. The experiment was therefore repeated with much greater refinements of detail by Michelson and Morley in 1887.

Fig. 3 shows the essentials of the apparatus employed, consisting of a metallic framework with the two arms DB and DC at right angles, and as nearly as possible equal in length. Metallic mirrors were fixed at B and C, perpendicular to BD and CD respectively. A mirror of unsilvered glass was fixed at D, making an angle of 45° with the line AB; and a telescope, with its axis perpendicular to AB, was placed at T. The whole apparatus was firmly bolted to a block of stone, floating in mercury, in such a

way that AB and CD lay in a horizontal plane, and the whole apparatus could, when desired, be turned slowly and steadily about a vertical axis. A beam of light from an electric arc or other convenient source of light at A would be divided at D into two portions. One of these portions would pass through D to B, from which it would be reflected back to D, where it would be again divided, part passing through D back to A, and part being reflected at D in a direction parallel to the axis of the telescope T. The other portion would be reflected from D to C, and back to D, where it would be divided, part being reflected back to A, and part passing through D in a direction

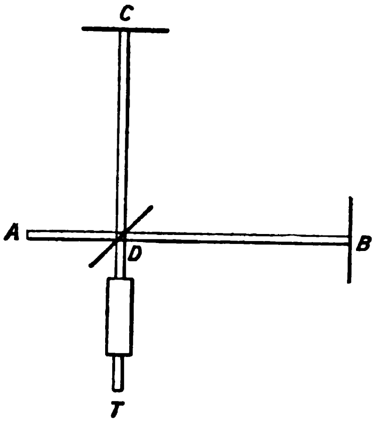


FIG. 3.

parallel to the axis of the telescope. Two beams of light from the source A would then enter the telescope, one taking the path ADBDT, and the other taking the path ADCDT. These two beams would give rise to a series of coloured interference fringes in the telescope if there were the least difference in the lengths of the two paths. Suppose the apparatus to be adjusted so that the fringes are as sharply defined as possible when the apparatus is so placed that AB represents the direction of the earth's orbital motion. The ray will then travel from D to B with a velocity $V+v$, and from B to D with a velocity $V-v$. Since the ratio of v^2 to V^2 is that of unity to one hundred millions, it can be shown by simple geometry[1] that the time occupied by the ray in travelling from D to B and back to D will be greater by one part in a hundred millions than when DB

[1] See *Æther and Matter*, p. 46.

is turned through a right angle, when the velocity will be unaffected by the earth's motion. In the apparatus employed in the 1887 experiments, the rays were reflected backwards and forwards by mirrors in such a way as to increase the effective lengths of the arms DB and DC to 11 metres each. Under these conditions there should, if the apparatus were moving with the earth through the ether at a speed of 30 kilometres a second, be a displacement of the fringes, amounting to two-fifths of the distance between two successive fringes, when the apparatus was turned through a right angle from its original position. The result was entirely negative, although the apparatus was sufficiently sensitive to have indicated a displacement of about a twentieth of the amount expected.

This negative result is of such fundamental importance that, before considering its bearings, and the explanation that has been given of it, it will be of interest to look a little further into the nature of the intervals of time which have to be measured, and how it is possible to make the measurements with sufficient accuracy. Since the effective length of DB in the experiment was 11 metres, the distance travelled by the light in its journey from D to B and back to D would be 22 metres if the system were at rest in the ether. Light would traverse this distance in

$$\frac{22}{300,000,000}\text{ sec.} = \frac{7\cdot\dot{3}}{100,000,000}\text{ sec.}$$

Now, we have mentioned that the earth's motion in the direction AB would increase this by one part in one hundred millions, so that the increase, to be determined by observation of the displacement of the fringes, would be

$$\frac{7\cdot\dot{3}}{100,000,000} \times \frac{1}{100,000,000}\text{ sec.} = \frac{7\cdot\dot{3}}{10,000,000,000,000,000}\text{ sec.},$$

that is to say, rather less than a thousand-billionth of a second.

To determine the limits of accuracy, let us suppose, for simplicity, that the source of light employed is homogeneous and emits only the light of the yellow line of the sodium spectrum. This is very near the middle of the visible spectrum, its wave-length being about six ten-thousandths of a millimetre, while the wave-lengths of the extreme visible blue rays and the extreme visible red rays are three ten-thousandths and eight ten-thousandths, respectively. The fringes would then be replaced by alternate bands of light and darkness, and the interval between two bands of maximum darkness (or maximum light) would correspond to the time occupied by light in travelling a distance equal to the wave-length of the light, viz. six ten-thousandths of a millimetre. The

observers, moreover, were able to read the displacements of the fringes with an error not much exceeding (if at all) a fiftieth of a fringe-width, so that the error in time, measured in seconds, would not much exceed

$$\frac{\text{Number of metres in one wave-length}}{\text{Velocity of light in metres per second}} \times \frac{1}{50}$$

$$= \frac{6}{10,000,000} \times \frac{1}{300,000,000} \times \frac{1}{50}$$

$$= \frac{1}{25,000,000,000,000,000} \text{ sec.}$$

This quantity, one twenty-five thousand billionth of a second, representing the smallest change in time that could be detected in the observations, is almost exactly a nineteenth part of the expected increase as given above, and differs very little from the observers' estimate of about one-twentieth.

A variation of the experiment with still greater refinements was carried out by Michelson in 1897, the arrangement being adapted to determine whether any difference could be detected in two horizontal paths each about 200 feet long, and one of which was 50 feet vertically over the other. If the ether *were* at all dragged along by the earth, the drag would be greater at the lower level than at the higher one, and the author concluded from the absolutely negative result that, if any such carrying along of the ether did occur, it must extend to distances of something like the earth's diameter of 8000 miles—a practical impossibility.

In these later experiments trouble was at first experienced from temperature effects, to eliminate which it was found necessary to enclose the path of the rays in tubes exhausted to a hundredth part of an atmosphere. The whole path was therefore practically *in vacuo*, that is to say, in free ether, so that the negative result could not be due, as has been suggested, to a modification of the ether by the atmosphere.

The earlier form of the Michelson-Morley experiment, as carried out in 1887, would have been accepted without question as demonstrating that the ether is carried along with the earth, had it not been for the apparently insuperable difficulties with which such a conclusion would have confronted us, as we have seen earlier in the present chapter. Some physicists did so accept it. The results of the 1897 experiment would, however, add still further to the difficulties involved in such a conclusion.

It was pointed out at about the same time by Fitzgerald and Lorentz that the null result of this experiment could be accounted for without assuming the ether to be carried along by the earth,

by supposing that the whole apparatus, and therefore, of course, all other solid bodies, undergo a certain contraction in the direction of motion when moving through the ether. When this hypothesis was first propounded it was regarded by many physicists as of too extraordinary a character to be taken seriously. Lorentz supported his suggestion by the arguments contained in the following paragraph[1]:—

"However extraordinary this hypothesis may appear at first sight, it must be admitted that it is by no means gratuitous, if we assume that the intermolecular forces act through the mediation of the ether in a manner similar to that which we know to be the case in regard to electric and magnetic forces. If that is so, the translation of the matter will most likely alter the action between two molecules or atoms in a manner similar to that in which it alters the attraction or repulsion between electrically charged particles. As then the form and the dimensions of a solid body are determined in the last resort by the intensity of the molecular forces, an alteration of the dimensions cannot well be left out of consideration. In its theoretical aspect there is thus nothing to be urged against the hypothesis. As regards its experimental aspect, we at once notice that the elongation or contraction which it implies is extraordinarily minute. It would involve a shortening in the diameter of the earth of about $6\frac{1}{2}$ centimetres. The only experimental arrangements in which it could come into evidence would be just of the type of this one of Michelson's which first suggested it."

Lorentz further showed that the necessary contraction, in the ratio of unity to $1-\frac{1}{2}v^2/V^2$, amounting in the case of bodies carried through the ether by the earth's orbital motion to one part in two hundred millions, could be deduced from what appeared to be reasonable assumptions with respect to the intermolecular forces. Larmor has shown that this law of contraction is necessarily true if the intermolecular forces of material bodies are entirely due to electrical actions between systems of electrons constituting the molecules. This will be more fully considered in Chapter IX.

The quantity which we have called V, the velocity of light in free ether, has the same value for all types of radiation, and may be called the radiation constant. When it is desired to emphasise this aspect, it will be denoted by the letter c, so that the expression given above for the shortening of bodies moving through the ether will be written $1-\frac{1}{2}v^2/c^2$.

[1] Translated by Sir Joseph Larmor from H. A. Lorentz's *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern*, Leiden, 1895, and reproduced in *Æther and Matter*, p. 185.

CHAPTER VII.

CONDUCTION IN GASES AND DIELECTRICS.

It is a noteworthy fact that, although the passage of the electric current through metals is very much the most familiar case of electric conduction, our knowledge of the mechanism by which that conduction is effected is of the most meagre and indefinite character in comparison with what is known about the methods of electric transmission through gases and liquids. Sir J. J. Thomson observes that this is no doubt partly due to the fact that the kinetic theory of gases provides us with a very much more definite conception of the structure of a gas than we at present possess of that of a metal. He suggests further that it may be due in part to the fact that the phenomena of metallic conduction are much less varied and show far fewer peculiarities than those which are observed in the passage of electricity through gases, and that it is these peculiarities which give a clue to the mechanism by which the transmission of the electric current is effected. It is for this reason that I am dealing with the phenomena of gaseous conduction at so early a stage.

The conductivity of gases in the normal state is so small that the slight leakage of electricity observed when a charged conductor is immersed in a gas was for a long time attributed entirely to the defective insulating power of the supports, or to the presence of particles of dust in the gas.

A long series of open-air measurements published by Elster and Geitel in 1900 showed that the rate of leakage varies greatly with time and place. It increases, for example, with the altitude, and is much greater in bright weather than in mist or fog. Shortly afterwards, both Geitel and Wilson found the rate of leakage from the leaves of an electroscope to be smaller in a closed vessel than in the open, and that it increases with the size of the vessel. They also found that, when the potential difference between the gold leaves and the walls of the containing vessel is increased, the rate of leakage increases only up to a fixed limit, until the potential difference becomes sufficient for sparking to

take place. This maximum rate of leakage Wilson found to be approximately proportional to the pressure of the gas.

Elster and Geitel also found the rate of leakage to be exceptionally high in cellars and caves filled with stagnant air. In one cave the leakage was no less than seven times as great as in the outside air.

There are many methods by which gases may be transformed from the condition of being almost non-conductors into one in which they may have a quite considerable conductivity. Among these are: raising the temperature of the gas beyond a certain point; or exposing it to the action of a sufficiently strong electric field, or to that of certain forms of radiation.

When a gas has been transformed into the conducting state, and the transforming agent is removed, the gas, under ordinary circumstances, returns only slowly and gradually to the normal condition; but this latter change may be greatly accelerated by filtering it through glass-wool, causing it to bubble through water, or subjecting it to a strong electric field, giving rise to a current through the gas.

These results appear to show that the conductivity is due to the presence of particles differing in some manner from the normal particles of the gas. The removal of the conductivity electrically shows that these particles must carry electric charges, and move under the influence of the electric field; and the fact that no electrification of the gas as a whole takes place as the result of this removal, shows that the charges must be both positive and negative. Owing to their analogy with the ions occurring in the electrolysis of solutions, the charged particles have also been called ions by Sir J. J. Thomson and others, and the process of transforming a gas into the conducting state by the production of these ions is now generally known as ionisation. It is found, however, that these gaseous ions differ greatly, both in their masses and in their electric charges, from the ions encountered in the electrolysis of solutions.

If two parallel metal plates be immersed in a gas in the conducting state, and a difference of potential established between them, an electric current will pass between the plates. The relation between the current and the potential difference may be investigated by connecting one of the parallel plates with one of the pairs of quadrants of an electrometer, the other pair being earthed. The other plate is connected to one terminal of a secondary battery, while the other terminal is earthed. The two pairs of electrometer quadrants are connected together until the observer is ready to take readings. The connection is then broken, and the rate of increase of the electrometer deflection

measures the current passing between the plates. With a small potential difference between the plates, it will be found that the current increases proportionately to this difference, which is in accordance with Ohm's law, as in the case both of ordinary metallic, and electrolytic, conduction.

It will be found, however, that as the potential difference is gradually increased, the rate of increase in the current becomes less, until the current reaches a maximum value known as the *saturation current*. A further increase in the potential difference, beyond the amount required to produce the saturation current, will produce no further increase in the current, until the electric field is strong enough to itself give rise to ionisation, after which the current will increase more rapidly than the potential difference. For a given distance between the plates, the potential difference required to produce ionisation is found to be directly proportional to the pressure of the gas.

With a constant difference of potential, more than sufficient to produce saturation, but not enough to cause ionisation, it will be found that an increase in the distance between the plates will give rise to an increase in the current, instead of a decrease, which would be observed if they were separated by a liquid or a solid.

This phenomenon can be very simply explained on the ionisation hypothesis. Let us suppose the ionising agent to produce ionisation uniformly throughout the gas, causing the ionisation to increase at the rate of $2n$ ions per unit of volume per second. Then, if A is the area of one of the plates, and d is their distance apart, the volume of gas included between them will be Ad, and the total ionisation per second will therefore be $2nAd$. If the movement of the ions under the action of the electric field were so rapid that each ion might be considered as reaching one or other plate the instant it was produced, nAd positive ions would be driven against the negative electrode, and nAd negative ions against the positive electrode, in a second, so that if e is the numerical value of the charge on each ion, and C the current, we should have

$$C = nAde.$$

The current would therefore, under these circumstances, at once become proportional to the volume Ad of the gas, and therefore if d were alone varied, the current would vary in the same ratio. The mean velocities of the positive and negative ions are proportional to the electric force, and are found to vary, under an E.M.F. of a volt per centimetre, from ·78 centimetre a second in the case of carbonic acid, to 7·2 centimetres a second in the

case of dry hydrogen. In consequence of these moderate velocities a small interval of time must elapse before the establishment of a steady state, but as soon as this is attained, the number of ions taken out of the gas in a given time must be equal to the number produced in the same time, and the current will therefore attain its maximum, or saturation, value. This direct proportionality between the saturation current and the distance between the plates was experimentally demonstrated in 1896 by Sir J. J. Thomson and E. Rutherford in the case of gases exposed to Röntgen rays.

Even in the absence of an electric force acting on the gas and removing some or all of the ions, the ionisation would never become complete, as the production of ions, by whatever ionising agent is employed, is accompanied by a continuous decay of ionisation, due to the recombination of positive and negative ions as the result of collisions. When the rate of ionisation balances the rate of decay, a steady condition will be maintained. It is clear, therefore, that the ionisation must begin to decrease as soon as the ionising agent ceases to act.

Another cause of decay in the ionisation of a gas is the diffusion of the ions through the gas until they come in contact with the walls of the containing vessel, to which they adhere, owing to the attraction caused by their electric charges. Townsend and others have shown that the rates of diffusion of the ions differ for different gases, and that they are considerably smaller for the gaseous ions than for the un-ionised particles of the same gases. It can easily be shown that the speeds of the various ions in an electric field must be proportional to their rates of diffusion. Zeleney found in 1898, from measurements of the speeds of various ions in the electric field, that the negative ions move more rapidly than the positive ones. This explains the fact that, under certain circumstances, ionised gases, originally unelectrified, may become positively charged. This occurs, for example, when such gases are blown through metal tubes, and is evidently due to more negative than positive ions being lost by diffusion to the sides of the tube.

When a charged ion moves in a field of magnetic force, it is deflected, its tendency being to move across the lines of magnetic force. If v is the velocity of the ion or charged particle in centimetres a second, e its electric charge, H the numerical value of the magnetic force at the position of the particle (e and H being measured in electromagnetic units), and θ the angle between the directions of H and v, then the force acting on it will be (see p. 112) $Hev \sin \theta$, and its direction will be at right angles both to H and v. The simplest case possible will be that of a particle moving in a

vacuum in a uniform magnetic field. The lines of force are then parallel straight lines. If no electric forces act upon the ion, the force acting upon it being only that due to the magnetic field, which is always at right angles to the path of the ion, the velocity of the latter will be constant in magnitude and direction. Moreover, the force on the ion being at right angles to the magnetic force, there will be no acceleration parallel to the latter, so that, in a uniform magnetic field, the velocity of the ion resolved parallel to the magnetic force will be constant, and therefore the direction in which the ion is moving will make a constant angle with that of the magnetic force. Let m be the mass of the ion in grammes, and r the radius of curvature of its path, then, in any magnetic field, the force in the direction of the normal will be $\frac{mv^2}{r}$, so that we shall have

$$\frac{mv^2}{r} = \mathrm{H}ev \sin \theta,$$

whence

$$r = \frac{mv}{e\mathrm{H} \sin \theta}.$$

Since v and θ are constant when no electric force is acting upon the ion, other than that arising from a uniform magnetic field, it follows that r is constant, and therefore the path of the particle will be a helix wound on a circular cylinder having its axis parallel to the lines of magnetic force. The diameter of this cylinder will be given by the left-hand member of the equation [1]

$$2r \sin^2 \theta = \frac{2mv \sin \theta}{e\mathrm{H}}.$$

The expression for the force shows that it changes sign, or in other words becomes reversed in direction, if the sign of either e or H is changed—that is to say, if the charge is changed from a positive to a negative one, or *vice versa*, or if the magnetic force is reversed in direction, but not if both these changes are made simultaneously. The result of such a reversal in the direction of the force will evidently be to reverse the direction in which the particle revolves round the line of force, or, in other words, to change a right-handed helix into a left-handed one, or *vice versa*; but it will not affect either the magnitude or the direction of the velocity parallel to the lines of force. In the special case of a particle projected at right angles to the lines of force, the helix will shrink into a circle whose diameter is $\frac{2mv}{e\mathrm{H}}$, so that the

[1] Sir G. Stokes, *Roy. Soc. Proc.*, March 1876.

particle will return to the point from which it was originally projected.

When the magnetic field is not uniform, so that the lines are curved, instead of being parallel straight lines, the consideration of various simple cases shows that the motion of the ion or charged particle will still tend to follow the lines of force in curves of a helical character, but in addition to this there will be, in some cases, a general drift of the particles.

In the year 1897 Sir J. J. Thomson succeeded in measuring the ratio of the charge of an ion to its mass in the case of the negatively charged cathode ray particles (see p. 107). By measuring the deflection of the particles in a uniform magnetic field with its lines of force at right angles to the paths of the ions, he determined, in the first place, the value of the ratio e/mv. In order to determine v, two metal plates were fixed in the tube parallel both to the lines of magnetic force and to the undeflected path of the particles, and the plates were maintained at a known difference of potential by connecting them to the terminals of a battery. An electric field was produced in this manner with its lines of force at right angles both to the lines of magnetic force and to the undisturbed direction of motion of the ions. The electric force of numerical value E due to this field tended to deflect the ion with a force $\mathrm{E}e$, acting in the same straight line as the electric force $\mathrm{H}ev$ due to the magnetic field. The plates were connected to the battery in such a way that these two forces acted in opposite directions, and one of the two fields was altered until they just balanced each other, as was shown by there being no deflection. This gave

$$\mathrm{E}e = \mathrm{H}ev$$

or

$$v = \frac{\mathrm{E}}{\mathrm{H}}.$$

This determined v, and therefore also the ratio e/m, since e/mv was already known.

When the tube was very highly exhausted, so as to practically eliminate the effect of the viscosity of the gas in resisting the motion of the ions, values of v were obtained as high as 60,000 miles a second—nearly a third of the velocity of light. In other experiments with less highly exhausted tubes values of v as low as 5000 miles a second were obtained, but in all cases the velocities were many thousand times the average velocity of hydrogen molecules at any available temperature, and therefore far greater than the velocity of any moving body hitherto known.

The value e/m is found to be quite independent of the velocity,

whatever the source of the particles, except in cases in which speeds approaching that of light are attained, when, as we shall see later, other considerations have to be taken into account. It was also quite independent of the nature of the gas in the tube and of the materials employed for the electrodes. The value of this ratio was found to be about 1700 times as great as in the case of the hydrogen ion in ordinary electrolysis.

The determination of the ratio for the positive ions is much more difficult, as the effect of the magnetic field in their case is very much smaller, and therefore very much stronger fields have to be employed. Measurements made by Wien and Sir J. J. Thomson on positive ions obtained by very different methods have shown that the ratio in their case is of the same order as those observed in liquid electrolysis.

The velocities of the negative ions in gases at very low pressure are much higher than those of the positive ones, in some cases attaining the speed of about half the velocity of light, and they do not appear to vary with the nature of the gas or of the electrodes; but these conditions do appear to affect the velocity of the far more slowly moving positively charged ions.

The question now arises whether the great excess of the ratio of charge to mass in the case of negatively charged gaseous ions over that found for liquid electrolysis is due to a difference in the charges, or in the masses, or in both. This question was also answered by Sir J. J. Thomson, who in 1898 and 1899 made determinations of the value of the negative charge of the gaseous ion, produced in the first case by Röntgen rays, and in the second by rays of ultra-violet light. The method was founded on the observation made by C. T. R. Wilson in 1897, that both positive and negative gaseous ions act as nuclei for the condensation of moisture, forming clouds, even in air quite free from dust.

The method consisted in forming a cloud in a dust-free gas containing free ions, by saturating it with water vapour, cooling it by a sudden slight expansion, and then observing the rate of fall of the drops through the gas. This observation, by the aid of a theorem due to Sir George Stokes, connecting the rate of fall of a drop of water with its radius and the viscosity of the gas, gave, the latter quantity being known, the volume of the drop. The quantity of water deposited from each cubic centimetre of gas was then determined by the aid of thermodynamic theory, and this gave the number of drops, and hence the number of ions per cubic centimetre. Now, knowing this latter quantity, and also the velocity of the ions under unit E.M.F., the charge of an ion could be deduced from the measurement of the current carried by these ions across each unit of area under the action of a given E.M.F.

The result obtained in each case was that the charge is the same as that of the hydrogen ion in the electrolysis of a liquid, from which it follows that the mass of the negatively charged gaseous ion is only about one seventeen-hundreth part of the mass of an atom of hydrogen, which was the smallest portion of matter hitherto recognised. The small differences observed between the velocities of the positive and negative ions in gases where the pressure is not extremely low would indicate that in such cases the negatively charged ions become attached to ordinary molecules of the gas.

A number of determinations of the ratio e/m, e being expressed in electrostatic instead of electromagnetic units, have given numerical values varying from $7{\cdot}6 \times 10^6$ to $1{\cdot}865 \times 10^7$, so that the ratio is a quantity of the order 10^7, m being expressed in C.G.S. units. Sir J. J. Thomson finds the most probable value of e to be about $3{\cdot}4 \times 10^{-10}$ electrostatic C.G.S. units. The charge on the hydrogen atom in ordinary electrolysis, calculated from experimental electrolytic data, by the aid of various assumptions as to the nature of molecules, in accordance with the principles of the kinetic theory of gases, is found to lie between the limits $1{\cdot}29 \times 10^{-10}$ and $6{\cdot}1 \times 10^{-10}$, so that the value of e may be identified with the charge carried by the atom of hydrogen in ordinary electrolysis. In electromagnetic units e is of the order 10^{-20}, its most probable value being $1{\cdot}2 \times 10^{-20}$ C.G.S. units.

The easiest method of obtaining a general grasp of the varied and interesting phenomena accompanying the discharge of electricity through gases is to classify them according to the method of ionisation employed. I shall therefore now proceed to consider the more important of these methods in detail.

Ionisation by Incandescent Solids.—As long ago as 1725 du Fay discovered that the air in the neighbourhood of red-hot metals is a conductor of electricity; and Becquerel noticed in 1853 that the conductivity was very largely increased when the metal was raised from a red to a white heat. In 1873 Professor Guthrie made the important discovery that there is a difference in this respect between positive and negative electrification, a red-hot ball being capable of retaining a negative charge, but not a positive one, while a white-hot ball can retain neither. In 1882 Elster and Geitel published the first of an elaborate series of researches on the ionisation produced by incandescent solids, and they and other observers found that the resulting electrification produced on neighbouring metallic plates depended on the nature and temperature of the hot wire employed, and also on the nature and the pressure of the gas surrounding it. It is found that

the ionisation is confined to a layer of gas immediately surrounding the hot solid.

Observations, made with platinum and carbon as the glowing substances, show that the formation of positive ions begins at a dull red heat, and that they increase in number as the temperature rises. Negative ions first make their appearance at a bright yellow heat, and their number increases more rapidly than that of the positive ones, until at very high temperatures the numbers of positive and negative ions are equal. The observed fact that the ratio of the charge on a negative ion to its mass is the same as in the cases considered earlier in this chapter shows that the same corpuscles, as Sir J. J. Thomson calls them (they are now more generally known as electrons), are the carriers. In the case of the positive ions, on the other hand, the carriers appear to consist mainly of molecules, or even small masses, emitted by disintegration of the incandescent substance, with the addition, in all probability, of molecules of occluded gases expelled from the interior.

Ionisation by Flames.—The fact that the gases from flames act as conductors was known as long ago as the time of Volta, who, in his experiments on contact electricity, made use of the fact to effect the discharge of the surface of an electrified non-conductor by passing it through a flame. Both high temperature and chemical combination are requisite for ionisation by means of flames. Thus, for example, no ionisation is produced by low-temperature flames, such as that of ether. The introduction of various metallic salts into a flame is found to cause a large increase in the ionisation.

Ionisation by Light.—The discovery of the effect of light, especially of the ultra-violet rays, in causing ionisation, originated in an observation made by Hertz in 1887, in the course of his classic researches on electric waves (see Chapter XIII.), that the incidence of ultra-violet light upon a spark gap facilitated the passage of the sparks. This led to a series of researches by Hallwachs, Hoor, Righi, and Stoletow, in which it was shown that a newly cleaned surface of zinc acquired a positive charge if subjected to ultra-violet radiation, the surrounding gas becoming at the same time negatively electrified. When the zinc was negatively charged in the first instance, this charge was rapidly lost; but when the charge was positive, no loss took place. The ultra-violet rays may be obtained from an arc lamp, from burning magnesium, or from induction coil sparks between terminals of zinc or cadmium. G. C. Schmidt and O. Knoblauch showed that quite a large number of substances exhibit similar photo-electric phenomena, and Elster and Geitel found that the more

electro-positive metals lose negative charges even in ordinary daylight. In the case of sodium and potassium, they found that the light of a petroleum lamp is effective in producing discharge, and that, in the case of the still more electro-positive metal rubidium, discharge is produced even by the light emitted by a glass rod heated to dull redness. It is noteworthy also that the light from a petroleum lamp is sufficient to cause a sensible increase in the conductivity of sodium vapour, which means that there is a sensible amount of ionisation.

Spark Discharge.—It was observed by Faraday that a greater potential difference is required to start the first spark than to maintain the discharge afterwards, and more recently Sir J. J. Thomson has shown that when a gas is very carefully dried it may be made to withstand a P.D. three or four times as great as that which is sufficient to produce a discharge under ordinary conditions; but when once a spark has passed, the P.D. falls to the value for moist gas, and cannot be raised to its former amount until the gas has been allowed to rest for some minutes. The greatest P.D. which can be applied to the terminals for an indefinite time without producing a spark is known as the *spark potential.* If a P.D. only just exceeding the spark potential is applied to the terminals, it may be as much as several minutes before a spark passes. This interval, which is called the *lag*, decreases as the P.D. increases. It is quite clear, as Jaumann pointed out in 1895, that during this interval some process must be going on in the gas which has the effect of transforming it into the conducting state. Thomson suggested that this transformation may be effected by means of ions which, under the influence of the electric field producing the spark, acquire a sufficient velocity to ionise the molecules of the gas, with which they come into collision. According to this view, the *lag* is the interval during which the ions and the current are continually increasing, until a steady state is reached, accompanied by the spark discharge, and it is clear that the *lag* will be diminished by any agent which increases the number of negative ions present in the gas, such, for example, as ultra-violet rays impinging on the negative electrode.

It is found experimentally that the spark potential is approximately proportional to the spark length, unless the latter is less than a certain value, which has been called the *critical spark length.* If the spark length falls below this critical point, which is found to be inversely proportional to the pressure of the gas, the spark potential rises for a time, and then falls rapidly as the spark length steadily decreases. If the spark length is maintained at a constant value well above the critical point, and the

pressure, starting with that of the atmosphere, is steadily decreased, the spark potential falls rapidly to a minimum value, and then begins to increase, but at a much slower rate than that of decrease.

Paschen, in 1889, arrived experimentally at the conclusion that the spark potential depends only on the product of the pressure of the gas and the spark length, *i.e.* on the mass of gas included between unit area of the electrodes. The same result can be arrived at, by the aid of considerations derived from the kinetic theory of gases, from Sir J. J. Thomson's theory of ionisation by impact.

The P.D. required to produce a spark of given length is in general independent of the material of the electrodes, though some observers have come to the conclusion that sparks pass more easily between aluminium electrodes (and perhaps this may apply also to magnesium) than between electrodes of other materials.

Discharge from Points.—The strength of the electric force tending to discharge the electrification from a charged conductor is greatest at the most pointed portion, and least at the most rounded portion, so that discharge will begin where the conductor is most pointed. The most convenient method of experimenting on the discharge from points is to make use of one very sharply pointed electrode, such as a needle, making the other electrode of considerable area—a plane surface, for example. Then, in the case of atmospheric pressure, luminous effects are only observed in the neighbourhood of the point, and the current through the other portions of the gas is carried entirely by ions of the same sign as that of the charge on the pointed electrode. If the pointed electrode is perpendicular to the plane one, it is found that, in order that electricity may stream from the point, its potential must differ from that of the plane by a certain amount, which Röntgen has called the *minimum potential.* The minimum potential depends upon the sharpness of the point, the pressure and nature of the gas, and the sign of the electrification of the point, being less when the point is negatively, than when it is positively, electrified. Warburg has shown that it is independent of the distance between the point and the plane, but the magnitude of the current which passes for a given P.D. diminishes rapidly as the distance between the point and the plane increases. Sir J. J. Thomson suggests that, when the electric field at the point reaches a certain intensity, a short spark probably passes from the pointed electrode to the air in its neighbourhood, and that both positive and negative ions are produced by this discharge along the line of its path, that those ions which have the

same sign as the electrification of the point are then driven from its neighbourhood into the surrounding gas, and ultimately, under the influence of the electric field, find their way to the metal plate to which the point is discharging. When these ions have attained a velocity proportional to the electric force acting upon them, the mechanical force which acts upon them will be transferred to the air through which they are moving, and so give rise to currents of air directed away from the point—a phenomenon which has long been known under the name of the electric wind. This forward motion of the air is accompanied by a reaction on the point tending to drive it backwards. Arrhenius showed in 1897 that for a given current, and a positively electrified point, the reaction is proportional to the pressure of the gas, and that, for different gases at the same pressure, it varies as the square root of the molecular weight of the gas. He found the reaction in the case of an equal current of negative electricity to be very much less, the difference increasing rapidly as the pressure of the gas was diminished; for example, at a pressure of 70 cm. of mercury, the reaction on the positive point was 1·9 times as great as that on the negative, while, when the pressure was reduced to 5·1 cm. it was 15 times as great. These results are clearly due to the relatively small mass of the negative ions compared with that of the positive ones, the effect of which on the velocity would necessarily increase rapidly with the diminution in the viscous resistance of the gas, caused by decreasing the pressure.

The Electric Arc.—If a circuit containing two pieces of carbon in contact with each other is traversed by a sufficiently strong current—which must be greater, the greater the diameter of the carbon electrodes—the resistance at the point of contact may become sufficient to raise the carbons to red, or even white heat. If they are then separated to a short distance, an arc of light will be formed between them; and if the source of electricity employed is sufficiently powerful to maintain a P.D. of 50 to 60 volts between the terminals, after their separation, this arc will be maintained, and in the meantime the carbon electrodes will gradually burn away.

The intense heat and light emitted by the electric arc make it of great industrial value for various purposes. The temperature of the arc itself is found to be considerably higher than that of either of the electrodes, and the temperature of the positive electrode is considerably higher than that of the negative one, the temperature of the former being, according to Violle, whose measurements were published in 1892, about 3500° C., and that of the latter about 2700° C. The extremity of the positive electrode soon becomes hollowed out into a crater-like shape,

while the negative one becomes pointed; and the positive terminal is found to lose weight much more rapidly than the negative one, so that, if the electrodes are formed of carbon rods of the same length and diameter, the positive one will burn away more quickly than the negative one. The temperature of the crater, when formed, is found to be constant, and is usually considered to be that at which carbon volatilises; the effect of increasing the current is merely to increase the size of the crater. When the current is increased beyond certain limits, there is found to be a drop of some 8 to 10 volts in the P.D. between the electrodes, and when this occurs a hissing sound is given out by the arc, which may develop into a roar when the increase of current is sufficiently great. Mrs Ayrton has shown

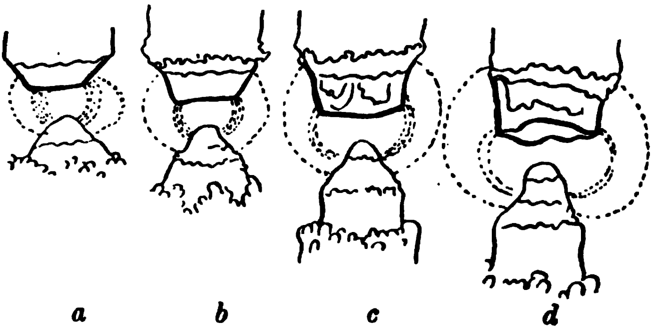


FIG. 4.

that the arc begins to hiss when the incandescence of the anode spreads beyond the crater, and extends up the sides of the carbon rod. Under these circumstances, the glowing carbon is no longer completely protected from oxidation by the carbon vapour, so that the incandescent portion becomes rapidly oxidised by the air. Mrs Ayrton found that, when the arc was formed in a closed vessel, the increase in current ceased to produce the hissing effect, as soon as the combustion of the oxygen in the vessel was complete, and that, on the introduction of fresh oxygen, hissing immediately recommenced. In fig. 4, which is from drawings made by Mrs Ayrton,[1] *a* and *b* illustrate the appearance of the electrodes with the quiet arc, and *c* and *d* with the hissing arc. Sir J. J. Thomson's theory of the arc discharge is that it is similar to that which takes place between terminals independently maintained in an incandescent condition. According to this view,

[1] Mrs Ayrton, *Journal Inst. Electrical Engineers*, xxviii., p. 400, 1899.

the bombardment of the cathode by the positive ions maintains its temperature at such a high value as to cause the continuous emission of negative ions from the cathode. These negative ions travel so much more rapidly than the positive ions, at the high temperature of the electric arc, that they carry by far the greater part of the discharge, and in addition to bombarding the anode, and so maintaining its incandescence, they ionise the carbon vapour emitted by the anode, either directly by collision, or indirectly by heating the anode, and in this way maintain the supply of positive ions which keep up the temperature of the cathode. According to this explanation, the essential feature in the discharge is the hot cathode, which has to supply the carriers of the greater portion of the discharge. It is necessary also for the anode to be maintained in an incandescent state, as otherwise it could not keep up the supply of positive ions which maintain the temperature of the cathode; but if the temperature of the cathode were maintained by other means, the current would continue to pass, even if the anode were allowed to become cold. Some experiments made by Professor Fleming in 1890 show this to be the case. By means of a third exploring carbon, forming, with one of the electrodes, an independent circuit containing a separate source of electricity, he showed that a current can be maintained between a hot cathode and a cold anode, but not between a cold cathode and a hot anode.

If the two electrodes are of different materials, the P.D. may depend upon the direction of the current; more particularly is this the case when one is a metal and the other carbon. Such an arc is maintained much more easily when the carbon is the negative terminal, and the metal the positive one, than when this is reversed. There is also a great difference among metals, with respect to the formation of the arc, when they are used as electrodes; thus, for example, in the case of brass, cadmium, and bismuth, the arc can only be maintained with great difficulty.

The arc can be deflected by means of a magnetic field in exactly the same manner as if it were a flexible conductor carrying a current. The result of such a deflection is evidently to increase the length of the arc, and just as the arc may be extinguished by increasing its length, so it can be blown out by means of a strong magnetic field.

Vacuum Tube Discharge.—When an electric discharge traverses a space occupied by gas at a very low pressure, the appearance of the discharge presents very clearly marked differences at different portions of its path, as is shown in fig. 5, which is a typical illustration of the discharge in what is commonly known as a

vacuum tube—*i.e.* a tube filled with highly rarefied gas, and furnished with terminal electrodes, A and C, between which the discharge passes. A thin luminous layer is observed over the surface of the cathode C, and extending beyond this along the tube is a comparatively dark space, which is usually known as the Crookes dark space. The extent of the latter increases as the pressure of the gas is decreased, and is also increased to a certain extent by increasing the current. The shape of the further boundary of this dark space is dependent on that of the cathode, coinciding approximately with the surfaces traced out by a series of perpendiculars of constant length to the surface of the cathode. Beyond the Crookes dark space there is a luminous region, known as the "negative glow," which is followed by a second dark space, generally known as the Faraday dark space. Extending from the further boundary of the Faraday dark space up to the anode, A, is another luminous region, known as the "positive column," which occupies by far the greater portion of the tube, except when this is a very short one. When the pressure and the current are within certain limits, this column breaks up into alternating dark and bright spaces, known as striæ, and which are indicated in the illustration.

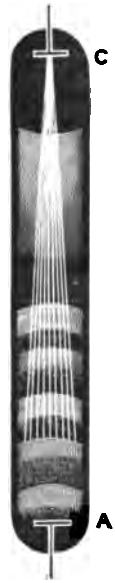


FIG. 5.

The electric force at any part of a vacuum tube during discharge may be measured by a pair of secondary electrodes, so arranged that they may be moved to different positions within the tube, always remaining at the same small fixed distance from each other, measured parallel to the length of the tube. Then, provided the conditions are such that there is a plentiful supply of both positive and negative ions in the vicinity of the secondary electrodes, these will rapidly assume the potential of the gas in contact with them, and the electric force can be determined by measuring their potential difference. It is found in this manner that the electric force is very large in the Crookes dark space, that it diminishes rapidly towards the negative glow, reaching a minimum value at a point which lies, either within the negative glow, or in the portion of the Faraday dark space just outside it. The electric force then increases until the positive column is reached. In the absence of striæ, it remains constant throughout the positive column in pure gases; but in impure or mixed gases there may be considerable variation, and the presence of striæ causes a series of fluctuations above and below the mean value

at any part of the column. Close to the anode a sudden drop of potential occurs, known as the anode fall of potential, giving rise to great electric force.

Sir J. J. Thomson has shown that these phenomena can be accounted for only by the electric field in the vicinity of the cathode itself giving rise to ionisation, and he suggests that this may be effected, either by the positive ions, in their rapid motion under the influence of the great electric force near the cathode, ionising the molecules of gas, or by these same positive ions striking against the surface of the cathode, and so communicating to the surface layer of negative ions sufficient energy to enable them to escape from the metal when it is raised to incandescence. He also suggests that the luminous glow which spreads over the surface of the cathode may very possibly be the seat of this ionisation. On either hypothesis as to the ionisation near the cathode, the result would be the starting of negative ions from the neighbourhood of the cathode, and these must necessarily be driven away from it by the strong electric force, and soon acquire velocities sufficient to ionise the gas through which they pass. This will produce a continuous supply of fresh positive ions, which will be attracted to the cathode, and there give rise to more negative ions.

These ions will, on account of the intensity of the electric field close to the cathode, leave its surface with a very high velocity, and, therefore, the number of collisions in a given length of path will be comparatively small, so that there will be but little resulting ionisation in the Crookes dark space. As the distance from the cathode increases, the ions, resulting from collisions between these primary ions and the molecules of the gas, will start with a much lower velocity, owing to the smaller electric force at their place of origin, and they will therefore give rise to more ionisation. The ionisation will, therefore, increase rapidly as the intensity of the field becomes less, and, assuming that luminosity accompanies ionisation, the dark space will be followed by a luminous interval, the negative glow.

After the electric force has fallen below a certain value, the ions will soon cease to retain sufficient energy to cause ionisation, so that there will be a second dark space, the Faraday dark space, and the current will be carried by the negative ions only in this portion of the tube. This will give a negative charge to the gas, and therefore cause the electric force to increase as the anode is approached, until ionisation is again produced, with its accompanying luminosity, and thus there will be alternations of dark and luminous spaces, producing the phenomenon of striation. In order that there may be no striation in the positive column, the

intensity of the field must be such that the number of ions produced in a given time is exactly equal to the number which re-enter into combination in the same time. The intensity of the field, as well as the amount of ionisation, and therefore the luminosity, will remain constant, giving rise to a uniform column. Where the electric force remains constant, there must be an equal number of positive and negative ions in a given volume of the gas, and therefore there must be a considerable number of positive ions produced in the neighbourhood of the cathode. Whether these are produced in the gas, or derived from the anode, it can be shown mathematically that there must be a finite drop in the potential at the anode, and it has already been pointed out that observation shows this to be the case.

Cathode Rays.—The name of cathode rays to designate the electric radiation from the cathode of a highly exhausted vacuum tube was first employed by Goldstein in 1876, who, however, then believed them to consist of ether waves. Their discovery may be considered to date from Plücker's observation, in 1859, of greenish-yellow phosphorescence close to the cathode of a highly exhausted tube. He found that the patches of phosphorescence changed their position under the influence of a magnet, but that the magnetic deflection differed in character from that of other portions of the discharge. Ten years later, Plücker's pupil, Hittorf, observed that a well-defined shadow could be obtained by interposing a solid body between a pointed cathode and the side of a vacuum tube. He also noted that the shadow was equally distinct, whether the interposed body were transparent or opaque, a good conductor or an insulator. Varley, as we have seen, suggested in 1871 that they really consisted of electrified particles projected from the cathode. Researches published in 1879, by Sir William Crookes, afforded very strong support to this view, but did not lead to its being generally accepted.

It was usually assumed at that time that, if the phenomena were due to electrically charged particles, they must consist of ordinary molecules of matter. Crookes's investigation showed that the impact of these rays on the walls of the tube, or upon interposed bodies, gave rise to heating and phosphorescence, and that they were deflected by a magnet exactly as if they were particles carrying electric charges. There were, however, numerous discrepancies between the behaviour of these particles and that of ordinary gaseous molecules, and Crookes, with a marvellous prescience, for which he received very little credit at the time, designated the conditions of their existence as a fourth state of matter. The observation made by Hertz, in 1892, that cathode rays were capable of penetrating thin gold-leaf or aluminium,

was generally considered as affording a strong argument against Crookes's theory. This objection, however, and others, fell to the ground with the discovery of the extremely small mass, about a seventeen-hundredth part of that of a molecule of hydrogen, which would have to be ascribed to these particles, which made it possible to account for their enormous velocities, and their great power of penetrating ordinary matter.

Lenard, a pupil of Hertz, succeeded in 1894 in constructing a tube with a window of thin aluminium foil in such a manner as to withstand the external atmospheric pressure when the tube was exhausted, and made a series of interesting observations on the cathode rays after their passage through this metallic window—whence such rays are often referred to as Lenard rays. These rays are found to be capable of producing phosphorescence by their impact on bodies outside the tube, and of affecting a photographic plate. In these and other ways the action of the Lenard rays, though invariably much more feeble than that of the more recently discovered Röntgen rays, is very similar in character to the action of the latter.

Canal Rays.—If the cathode in a vacuum tube is perforated with a number of holes, then, provided the pressure of the gas is within certain limits, luminous streams or pencils of light may be seen passing through these holes and emerging on the side of the cathode which is further from the anode. These rays were discovered by Goldstein in 1886, who gave them the name of canal rays. They give rise to incandescence where they strike the glass. Wien, in the same year, succeeded in showing that these rays could be deflected both by magnetic and electric fields, though the deflections obtained were extremely minute in comparison with those produced by corresponding fields in the case of cathode rays. The direction of deflection showed that the canal rays consist of streams of positively electrified particles, and the smallness in its amount showed that the velocity of the particles is very much smaller than that of the cathode-ray particles. Wien also measured the ratio of the electric charge carried by a particle to its mass, and found it to be of the same order as in the ordinary electrolysis of solutions. Sir Joseph Thomson suggests that these rays are due simply to the passage through the perforation of portions of the stream of positively charged particles which by their impact on the cathode give rise to the emission of the cathode rays.

Sir J. J. Thomson has recently (*Phil. Mag.*, May 1909) made a series of determinations of the ratio e/m in various gases for the particles constituting the canal rays, or rays of positive electricity, as he prefers to call them. All the gases examined

showed the presence of two kinds of particles for which e/m has the maximum values, 10^4, corresponding to the atom of hydrogen, and 5×10^3, corresponding to the ordinary alpha particles; while in the case of helium a third kind of particle was observed for which $e/m = 2{\cdot}5 \times 10^3$. He points out that the agreement observed in the maximum value of e/m for different pressures shows that it is a true maximum, and excludes the possibility of there being other more deflected rays not strong enough to produce visible phosphorescence. For, if the value of e/m for a particle that had never temporarily lost its charge by collision were greater than 10^4, larger values for e/m should be observed at low pressures than at high ones.

Röntgen Rays.—Röntgen discovered, in 1895, that when a vacuum tube was so highly exhausted that the walls of the tube became vividly phosphorescent during the continuance of the discharge, rays which were propagated in straight lines were emitted into the space outside the tube. He found that a plate covered with a phosphorescent substance, such as potassium platino-cyanide, became luminous when interposed in the path of these rays, and that the interposition of various solids between the tube and the plate gave rise to shadows of greatly varying intensity, the depth and sharpness of the shadow being greater, as a general rule, the greater the density of the interposed substance; thus, for example, when a hand was placed between the tube and the plate, the outlines of the bones became distinctly visible in the shadows cast upon the screen, owing to the bones casting much stronger shadows than the surrounding flesh. When a closed wooden box containing pieces of metal was employed, the outlines of the metal were distinctly shown, the metal casting a very deep shadow, and the wood only a comparatively slight one.

The rays were also found to affect a photographic plate, so that shadow photographs could easily be obtained. These photographs are now extensively used for locating bullets and other foreign materials within the body, and for studying the conditious of bone fractures and various derangements of the internal organs.

Increasing the exhaustion of the tube, with the requisite corresponding increase in the potential difference between its electrodes, greatly increases the penetrative power, now generally called the "hardness," of Röntgen rays. This may be very well shown by gradually increasing the exhaustion of the tube while a hand is interposed between it and a phosphorescent screen. As the pressure of the gas is gradually reduced, the first production of rays will be indicated by the appearance of faint indications of the bones, these first rays being so "soft" that they are stopped by the flesh as well as by the bones. As the exhaustion proceeds,

the outlines of the bones are seen to become more and more distinct; and then, as the hardness of the rays increases still further, the distinctness begins to diminish, as the rays are now able to penetrate the bones as well as the flesh. The hardness of the rays depends, not only on the exhaustion, but also on other circumstances not yet fully determined; and not only are differences found to exist in the hardness of rays emitted from different tubes, but the same bulb is often found to emit, at the same time, rays which vary considerably in hardness.

Röntgen rays have an ionising effect, which varies to a large extent for different gases; and, in the case of any one gas, the production of ions in a given volume is found to be proportional to the pressure. The source of these rays is found to be the point where the cathode rays strike the tube, or in the case of a *focus tube*—viz. one with a hollow cathode, and having a plate made of some infusible substance placed at the point of convergence of the cathode rays—the part of the plate on which they impinge. Similarly, when Röntgen rays impinge upon a solid, a liquid, or even a gas, secondary Röntgen rays are generated. In the case of impact against a solid or a liquid, they are of a much less penetrating character than the primary ones. Röntgen rays do not undergo refraction in passing from one medium to another, and there is no evidence of their undergoing any deflection when traversing magnetic fields of sufficient intensity to give rise to very large deflections of cathode rays. It is evident from these considerations that Röntgen rays do not consist of streams of electrically charged particles, and there are sound reasons, as will be shown later, for believing that they are pulses in the ether produced by the sudden stoppage of cathode ray particles. We may consider them, in fact, to be in many respects similar to visible light-waves, consisting, like the latter, of electric and magnetic forces at right angles to each other and to the direction of propagation. The Röntgen ray pulses, however, differ from those of ordinary light in their want of any regular periodic character, and in the extremely small thickness of the pulses compared with those of ordinary light. Sir George Stokes has shown that the small thickness of the pulse is sufficient to account for their not being capable of refraction, and for other differences between Röntgen rays and ordinary light-waves.[1]

Conduction in Dielectrics.—Such electric conductivity as is exhibited by dielectrics was supposed, until quite recently, to be mainly, if not entirely, of an electrolytic nature, but the latest investigations appear to indicate that the process of conduction in dielectrics resembles the corresponding phenomenon in dense

[1] See Chapter XIII.

gases far more closely than its course in electrolytes. The basis for this new point of view was provided by P. Curie's discovery[1] that the conductivity of dielectrics is increased by exposure to radium radiation, and that this conductivity exhibits a departure from Ohm's law, which naturally suggested the ionising action of the radiation as the cause of the increase. Further investigations by Bécquerel[2] and Becker[3] confirmed these conclusions, and the phenomena have been recently made the subject of a very thorough investigation by Jaffé,[4] who finds that the current in dielectrics exposed to radium radiation may be divided into two parts, one of which obeys Ohm's law, and an ionisation current which does not follow Ohm's law.

The former current, which is probably of an electrolytic nature, appeared to be due entirely to the presence of impurities. Jaffé found it impossible, in spite of the most elaborate precautions, to eliminate the impurities entirely; but in the purest specimens obtainable the current obeying Ohm's law was reduced to an exceedingly small percentage of the whole. The ionisation current itself appeared to be capable of resolution into two distinct portions, one arising from mobile ions similar to those existing in ionised gases, and which was found to attain the condition of *saturation* in a field of a thousand volts per centimetre; the other portion showed no signs of saturation in a field of seven thousand volts per centimetre, and was ascribed by Jaffé to slowly moving ions, apparently of the nature of electrolytic ions.

These conclusions are confirmed by experiments made by B. Hodgson[5] on heavy paraffin oil, solid paraffin, glass, ebonite, benzene, toluene, aniline, and carbon disulphide, of the highest attainable purity, all specimens being rejected in which the ordinary conduction current was in excess of 10^{-13} ampere in an applied field of about a hundred volts per centimetre. The conduction current was first determined before subjecting the dielectric to the radium radiation, and the ionisation current was determined by subtracting this from the total current observed under the influence of the radiation. When the radium was applied, a considerable immediate increase in conductivity was invariably observed, and when it was allowed to act continuously, the current was found to increase steadily with the time. On withdrawing the dielectric from the influence of the radium, the

[1] *Comptes Rendus*, vol. cxxxiv., 1902, p. 420.
[2] *Ibid.*, vol. cxxxvi., 1904, p. 1173.
[3] *Annalen der Physik*, vol. xiii., 1904, p. 394.
[4] *Ibid.*, vol. xxv., 1908, p. 257.
[5] *Phil. Mag.*, vol. xviii., 1909, p. 252.

current was always found to diminish, and ultimately to fall to its initial small value. This complete recovery of its original insulating power by the dielectric occupied a considerable time—as much as thirteen hours in the case of one specimen of paraffin.

Some experiments were also made upon ordinary commercial vaseline, which was found to be rather irregular in its action. Its conductivity proper decreased very rapidly, so much so as to mask completely the increase due to the radiation, unless sufficient time were allowed to elapse for its own minimum conductivity to be attained before exposure to the radiation.

Note to p. 94.—It follows from the definition of the electromagnetic unit of current on page 26 that the force exerted by an element ds of a current C, on a magnetic pole m at a distance r from ds, measured along a normal to the latter, will be represented by the expression $m\mathrm{C}ds/r^2$. If the current be due to an electron with the charge e, moving with the velocity v along the element ds, $\mathrm{C}ds$ will be replaced by ev, so that the magnetic force on the pole m, and therefore also the force exerted by the pole on the moving electron, will be numerically represented by the expression mev/r^2. Hence generally, the force on the electron will be $\mathrm{H}'ev$, where H′ is the component of magnetic force, at the place occupied by e, perpendicular to the direction of its motion. If this component H′ arise from a resultant magnetic force H making an angle θ with the direction of motion, H′ will have to be replaced by $\mathrm{H} \sin \theta$, so that the force acting on the electron will be $\mathrm{H}ev \sin \theta$, as stated on page 94. See also Foster and Porter's *Electricity and Magnetism*, third edition, page 473.

CHAPTER VIII.

THE FARADAY-MAXWELL THEORY.[1]

We saw in Chapter III. that the earlier theories of electric actions were based on the scheme of central forces, acting at a distance between the particles of matter, which Newton and his successors had developed in order to correlate the observed phenomena of the motions of the heavenly bodies, in accordance with the law of gravitation which he had formulated. Hypothetical distributions of quasi-material substances, under the names of electrical and magnetic fluids, were assumed. Representations of observed electrical and magnetic phenomena were formed on the assumption of forces acting at a distance, like those assumed to account for gravitational phenomena, between the particles of these fluids, and these assumptions were found to be perfectly adequate for the representation of the phenomena then known. We saw, however, that Newton himself and a few others among the greater physicists remained unsatisfied with representations of actions between distant bodies without any intervening mechanism for their transmission. They could not believe, to quote Newton's own expression, "that matter can act where it is not."

When Cavendish's unpublished papers were unearthed by Clerk-Maxwell and published in 1879, it first became known that he had already, in the eighteenth century, discovered the fact that the nature of the dielectric medium separating a system of conductors plays an important part in the transmission of electric actions between the charged and other conductors. The fact was rediscovered by Faraday during the first half of the nineteenth century, and formed the starting-point of what is generally known as the Faraday-Maxwell theory, the now universally accepted basis for the representation of electric and magnetic phenomena.

It was mentioned in Chapter I. that Faraday was the first to realise the significance of the curves traced out by iron filings scattered in the neighbourhood of a magnet, although Gilbert had observed and called attention to them. If the mechanical actions

[1] See Appendix C.

observed between magnets or electrically charged bodies are transmitted by means of an intervening medium, this must necessarily be in a state of mechanical stress. Faraday concluded that the stresses indicated by magnetic and electrical phenomena were of the nature of tensions along the lines of force combined with pressures in all directions at right angles to them, the tension and the pressure at every point being numerically equal. Faraday pictured this state of the medium to himself by means of lines of force, which he regarded as having an actual concrete existence, like a bundle of stretched springs, in fact; and he showed that the phenomena of an electric or magnetic field might be explained on the assumption that a tension exists along the lines of force tending to make them become shorter, and that lines of force have a tendency to repel or to attract each other, according as the component of either parallel to the other is similarly or oppositely directed. The lines of electric force start from positively, and end on negatively, charged bodies, while the lines of magnetic force invariably form closed curves.

Maxwell considered this distribution of stress to be the only one consistent with the observed mechanical action, and also with the equilibrium of the fluid ether which surrounds them. He therefore starts by assuming the actual existence of this state of stress, and then proceeds to develop its consequences. His first step, following Faraday, was to ascribe all electrical phenomena to changes in polarisation of the dielectric. An elementary portion of any body is said to be polarised when it acquires equal and opposite properties on two opposite sides. Its properties will then be reversed on reversing the elements in such a manner as to interchange the position of these two sides. A permanent magnet is a familar example of such a body, the polarity being present in its smallest parts.

It is now recognised that the distribution of stress here considered corresponds to the actual system of forces called into action, only provided there is no resulting polarisation of the medium (see Appendix D). We shall see that this is the case in ether free from electrons, but not in any material medium. Maxwell's procedure, however, implies the existence of polarisation in the ether as well as in material substances, and therefore requires correction. The necessary modifications will be indicated as we proceed.

The first attempt at a mathematical theory of polarised bodies was Poisson's theory of magnetisation. Lord Kelvin's papers on "A Mathematical Theory of Magnetism," which were published in 1849, added but little to the actual results obtained by Poisson; but by avoiding all reference to molecules, and treating the media

as continuous, he was able to dispense with Poisson's artificial hypotheses, and to develop the theory of polarity on a purely dynamical basis. In these papers Lord Kelvin explained, for the first time, the nature of the dielectric action discovered by Cavendish and Faraday, and observed that, "however different physically, the positive laws of the phenomena of magnetic polarity and dielectric polarity are the same."

Faraday came to the conclusion, from his electrostatic experiments, that it is impossible to charge any portion of matter with an absolute and independent charge of electricity, as it appeared, from the results of a very large number of experiments, that whatever electrical actions took place between bodies surrounded by a metallic vessel, it was impossible to detect any change in the external electrification. This would not be the case if a portion of electricity could in any way be absorbed or come into existence in a body without being connected by lines of force with an equal portion of electricity of the opposite kind. He was therefore led to look upon induction as "the essential function both in the first development and the consequent phenomena of electricity," and he defined induction as consisting of a polarised state of the particles of the dielectric, each particle being positive on one side and negative on the other, the positive and the negative electrification of each particle being always exactly equal.

We now have to consider the concept of "electric displacement," which has always been one of the difficulties of Maxwell's theory. It was introduced to make the dielectric polarisation satisfy the conditions necessary in order that all electric currents should flow in streams round complete circuits, which would not only simplify the theoretical development, but appears physically probable. Electrostatic experiments described in Chapter I. show that, when a charged body is introduced into a closed metallic vessel without touching it, an equal induced charge of the same kind appears on the outside of the vessel, and this appears to indicate that no electricity can be introduced into a closed space without producing an outflow of exactly equal amount. That is to say, if electricity is to be treated as a fluid, it must be considered as incompressible, so that the quantity within a given space cannot be either increased or diminished.

The dielectric polarisation is of the kind which has already been referred to as a vector quantity, that is to say, a quantity which has direction as well as magnitude, and is such that a reversal of its direction reverses its sign. The name vector is derived from the line drawn from the original to the final position of a moving point, and which therefore represents its displacement. Such quantities are added together, like velocities and forces, by the

rule most commonly known as the parallelogram of forces. A rotation about an axis is a vector quantity of similar character, being capable of complete specification in terms of a measured length along the axis, that is to say, by a straight line.

Another class of vector quantities are what Maxwell calls fluxes, which are defined with reference to an area. The flow of heat in any direction at any point in a solid body is an example of a flux, being defined as the quantity of heat which crosses a small area drawn perpendicular to that direction divided by the area and by the time. Magnetic force and electric force belong to the first class of vector quantities, while the corresponding fluxes, magnetic induction, electric induction, and electric current, belong to the second.[1]

Vector quantities, such as the electric force at a point will be represented by block capitals, as **E**, and the corresponding Roman capitals, as E, will be employed when we are considering only the numerical values of the vectors, without reference to their direction.

It should be noted that there are other more complex physical quantities which are related to directions in space, and which are not simply vectors. Stresses (and the resulting strains) are quantities of this class, and they require nine numerical quantities for their specification, while vectors only require three.[2] A flux like that of an incompressible fluid is called a *stream vector* throughout any region within which there is no gain or loss of fluid except at certain isolated points, the *sources*, where the fluid enters the region, and the *sinks*, where it leaves the region. Such a region can be divided up into tubes of flow, each of which may be considered as independent, with no leakage from one to the other. For this reason Maxwell calls a flux of this kind a *solenoidal* vector, and Lord Kelvin calls it a *circuital* vector. The flow of heat in a solid body is an example of a flux which is not, in general, a stream vector, as some portions of the body will be growing hotter and others colder, so that if the flow were represented by that of an incompressible fluid, every point would have to be considered as either a source or a sink. The numerical value of the quantity of heat appearing in, or disappearing from, any small element of volume in unit time, divided by the small volume, is called the *concentration* or the *convergence* of the vector

[1] Maxwell's division of vectors into two classes, according to whether they are defined in reference to a straight line or to an area, is a mere convenient classification according to the manner in which any given vector is defined. It does not necessarily involve any physical distinction, as many vectors may be defined in either way, and therefore placed in either class, at will.

[2] Stresses and strains are expressed mathematically as linear and vector functions of a vector.

at the point considered. When a steady condition is attained, so that the temperature at each point remains constant, the concentration will become null, and the flux will become a stream vector. When the concentration of a vector quantity is negative, it is called the *divergence* of the vector.

In order that electric currents due entirely to changes in the dielectric polarisation should always flow in complete circuits, this quantity must be a stream vector. Now, on the theories of Poisson and Lord Kelvin, the induced polarity does not possess this property, and Maxwell therefore assumed the existence of an altogether new effect, a flux which he called electric displacement, possessing the stream property everywhere except at certain isolated points in the medium where there was assumed to be positive or negative concentration of the displacement. The total amount of this concentration, summed up for the region surrounding each isolated point, he called the electric charge at that point. These charges were assumed to consist of permanent ether forms involving a condition of intrinsic strain, capable of free motion through the ether, but which could not be either created or destroyed by any physical process. They might very well be called *strain-forms*, and we shall see in the next chapter that Lorentz and Larmor have succeeded in making this concept a perfectly definite one, completely adapted for the development of the purely dynamical electrical theory which Maxwell has shown to be possible, as we saw in Chapter III. It is, moreover, a perfectly simple matter to form a mental picture of such a strain-form by the aid of the Kelvin gyroscopic model of the ether. We saw in Chapter I. that the existence of discrete electric charges or electrons is indicated by Faraday's laws of electrolysis, and in Chapter VII. we obtained definite experimental evidence that the negative electrons can be isolated from, and exist independently of, ordinary gravitating matter.

I cannot do better than quote in full the paragraphs in which Maxwell leads up to and introduces the concept of electric displacement (treatise, article 60):—

"From the hypothesis that electric action is not a direct action between bodies at a distance, but is exerted by means of the medium between the bodies, we have deduced that this medium must be in a state of stress. We have also ascertained the character of the stress, and compared it with the stresses which may occur in solid bodies. Along the lines of force there is tension, and perpendicular to them there is pressure, the numerical magnitude of these forces being equal,[1] and each proportional

[1] This equality, as pointed out on p. 122, only holds good in ether free from electrons, and does not therefore, as assumed by Maxwell, apply to the case of a material dielectric.

to the square of the resultant force at the point. Having established these results, we are prepared to take another step, and to form an idea of the nature of the electric polarisation of the dielectric medium.

"An elementary portion of a body may be said to be polarised when it acquires equal and opposite properties on two opposite sides. The idea of internal polarity may be studied to the greatest advantage as exemplified in permanent magnets, and it will be explained at greater length when we come to treat of magnetism.

"The electric polarisation of an elementary portion of a dielectric is a forced state into which the medium is thrown by the action of electromotive force,[1] and which disappears when that force is removed. We may conceive it to consist in what we may call an electrical displacement, produced by the electromotive force. When the electromotive force acts on a conducting medium it produces a current through it, but if the medium is a non-conductor or dielectric, the current cannot flow through the medium, but the electricity is displaced within the medium in the direction of the electromotive force, the extent of this displacement depending on the magnitude of the electromotive force, so that if the electromotive force increases or diminishes the electric displacement increases or diminishes in the same ratio."

"The amount of the displacement is measured by the quantity of electricity which crosses unit area, while the displacement increases from zero to its actual amount. This, therefore, is the measure of the electric polarisation.

"The analogy between the action of electromotive force in producing electric displacement and of ordinary mechanical force in producing the displacement of an elastic body is so obvious that I have ventured to call the ratio of the electromotive force to the corresponding electric displacement the *coefficient of electric elasticity* of the medium. This coefficient is different in different media, and varies inversely as the specific inductive capacity of each medium.

"The variations of electric displacement evidently constitute electric currents. These currents, however, can only exist during the variation of the displacement; and therefore since the displacement cannot exceed a certain value without causing disruptive discharge, they cannot be continued indefinitely in the same direction, like the currents through conductors."

Now, let us suppose a charge of electricity, e, to be communicated to a spherical conductor having its centre at the point O (fig. 6), and imagine a concentric spherical surface of radius r to be described, enclosing the spherical conductor. Let V be the potential of the conductor, supposed to be due entirely to the presence of the charge e, then the charge will be uniformly distributed over the surface of the conductor, and if the charging

[1] It would, I think, be preferable to employ the term electric force here, and to confine the term electromotive force to the result of the summation of the electric force along a line (see p. 9). Maxwell's indiscriminate use of the term both in this sense and in that of the electric force at a point is confusing to the student.

is effected by a uniform inflow at every part of the surrounding spherical surface, the flow of electricity across unit area of this surface will be the quotient of the charge e by the area $4\pi r^2$ of the outer sphere.

According to Maxwell's theory, this must be accompanied by an outward displacement through the spherical surface, the total amount D of which will be proportional to the product of the resultant force at the surface of the sphere into the area of the surface. Now, the charge e being uniformly distributed over the surface of the spherical conductor, it can easily be shown that the force it produces at any external point will be the same as if the charge were concentrated at the centre O.[1] If the surrounding dielectric is air, the electric force $\mathbf{E}$ (in electrostatic units) at any point P of the outer sphere will therefore have the numerical value e/r^2, so that the total displacement will be proportional to the product of e/r^2 by $4\pi r^2$, viz. to e.

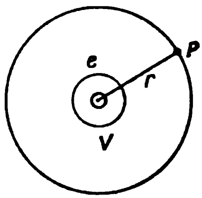


FIG. 6.

Therefore the whole displacement D is proportional to the charge e, and is independent of the radius r of the spherical surface.

Now let the displacement be increased by the small amount δD from D to $D + \delta D$, and consider the work done on the medium in a region included between two spherical surfaces concentric with O, the potential of the inner one being V_1, and that of the outer one V_2. Let d be the difference in the radii of the two surfaces, and let the numerical values of the electric force be E_1 on the inner one, and E_2 on the outer one. Then the work done, on the medium within this region, in augmenting D will be

$$(E_1 - E_2)d\delta D = (V_1 - V_2)\delta D.$$

If we now make the inner surface coincide with that of the spherical conductor, and the radius of the other infinite, V_1 will become equal to V, and V_2 will be zero, so that the whole work done in the surrounding medium will be $V\delta D$. But, in the

[1] See Foster and Porter's *Electricity and Magnetism*, p. 26.

ordinary theory, the work done in increasing the charge by the small amount δe is $V\delta e$. It follows, therefore, that the increase in the displacement is equal to the increase in the charge, and as the displacement has been shown to be proportional to the charge, there will be no displacement when there is no charge, so that the total displacement across any spherical surface enclosing and concentric with the surface of the spherical conductor must be equal to the charge. This equality will presently be shown to hold good for any finite closed surface, and not only for a concentric spherical one (see p. 133).

The electric displacement through unit area of the spherical surface P, measured in electrostatic units, will therefore have the numerical value $e/4\pi r^2$. If the dielectric is air, then $e/r^2 = E$, so that the displacement will be $\mathsf{E}/4\pi$, measured in electrostatic units.

If the air is replaced by a dielectric of specific inductive capacity K, the electric force will be decreased in the ratio K to 1, so that the general expression for the electric displacement D across unit area perpendicular to E, measured in electrostatic units, will be

$$\mathsf{D} = \frac{K\mathsf{E}}{4\pi}.$$

If D, E, and K are all expressed in electromagnetic units, the same expression will hold good, but, when K is expressed in electromagnetic units, its value for air is $1/c^2$, where c is the radiation constant of the ether, so that the displacement in air expressed in electromagnetic units will be

$$\mathsf{D} = \frac{\mathsf{E}}{4\pi c^2}.$$

As the quantity K is usually tabulated in electrostatic units, it is convenient to retain the numerical values for K so obtained, so that in formulæ expressed in electromagnetic units, which are the ones universally employed, K will have to be replaced by K/c^2, instead of including the factor $1/c^2$ in a new electromagnetic constant (K). The general expression for the electric displacement expressed in electromagnetic measure will therefore be

$$\mathsf{D} = \frac{(K)\mathsf{E}}{4\pi} = \frac{K\mathsf{E}}{4\pi c^2}.$$

Crystalline media are, in general, not isotropic, and in such cases K will no longer be a quantity having no reference to direction, or *scalar* quantity, as it is then termed, but will be a quantity

requiring, like a stress or a strain, nine numbers for its specification. In mathematical language, K will then be a linear and vector function of a vector.

The electric displacement across any area multiplied by 4π is called by Maxwell the induction across the area. The direction of induction will therefore be the same as that of the electric force whenever K is a scalar quantity. In crystalline media, however, where K varies with the direction, the direction of the induction will not generally coincide with that of the force. Faraday's so-called lines and tubes of electric force are really lines and tubes of induction, and we see from the above that the latter will be identical with lines and tubes of force, except in crystalline media. In these media it is therefore important to remember that the Faraday tubes are tubes of induction, and that their directions are not those of the lines of electric force. The formal definition of a Faraday tube will be given in the next chapter, but it may be noted in the meantime that the electric induction through any area is measured by the algebraic sum of the Faraday tubes which traverse it.

Since the introduction of an electric charge e within a closed spherical surface causes a simultaneous outward displacement e through the surface, we see that the introduction of the concept of electric displacement leads to the consequence that the motions of electricity invariably have the character of the flow of an incompressible fluid. We may therefore represent any electric flux by a flow of water, regarded as absolutely, instead of only approximately, incompressible. A simple representation of the introduction of an electric charge into any region may then be obtained by imagining a cylindrical tube containing a movable piston to be placed with one extremity within the region, and the other outside it, and displacing the piston in the direction which will cause a flow of water into the region. An equal amount will then be displaced outwards through the boundary of the region. The lines of flow will pass through the tube, leaving it at the end within the region, and returning to the other end. A series of pipes whose sides were traced out by lines of flow might then be made to complete the circuit of which the original tube formed a portion, and these, being tubes of flow, would not alter the circulation, but would merely serve to indicate the complete circuits formed by the flow. With pipes free from obstructions, a circulation would be maintained if the piston were replaced by a continuously acting pump maintaining a steady pressure. The conditions of a dielectric might be represented by introducing elastic diaphragms into the tubes, in which case there would be no continuous circulation unless the pressure became sufficient to

rupture the diaphragms, but every increase of pressure would increase the quantity of water introduced into the region, with an equal increase in the amount passing outwards through the boundary to complete the circuit.

The special features of this theory are enumerated as follows by Maxwell:—

"That the energy of electrification resides in the dielectric medium, whether that medium be solid, liquid, or gaseous, dense or rare, or even deprived of ordinary gross matter, provided it be still capable of transmitting electrical action.

"That the energy in any part of the medium is stored up in the form of a state of constraint called electric polarisation, the amount of which depends on the resultant electromotive force at the place.

"That electromotive force acting on a dielectric produces what we have called electric displacement, the relation between the force and the displacement being in the most general case of a kind to be afterwards investigated in treating of conduction, but in the most important case the force is in the same direction as the displacement, and is numerically equal to the displacement multiplied by a quantity which we have called the coefficient of electric elasticity of the dielectric.

"That the energy per unit of volume of the dielectric arising from the electric polarisation is half the product of the electromotive force and the electric displacement, multiplied, if necessary, by the cosine of the angle between their directions.

"That in fluid dielectrics the electric polarisation is accompanied by a tension in the direction of the lines of force combined with an equal[1] pressure in all directions at right angles to the lines of force, *the amount of the tension or pressure per unit of area being numerically equal to the energy per unit of volume at the same place.*

"That the surfaces of any elementary portion into which we may conceive the volume of the dielectric divided must be conceived to be electrified, so that the surface density[2] at any point of the surface is equal in magnitude to the displacement through that point of the surface *reckoned inwards*, so that if the displacement is in the positive direction, the surface of the element will be electrified negatively on the positive side and positively on the negative side. These superficial electrifications will in general destroy one another when consecutive elements are considered, except where the dielectric has an internal charge, or at the surface of the dielectric.

"That whatever electricity may be, and whatever we may understand

---

[1] This statement must be corrected by substituting the words "a constant pressure" for an equal pressure, and deleting the words printed in italics (see Appendix D). We shall see (p. 155) that the constant pressure at right angles to the lines of force is not, in material dielectrics, of equal numerical value to the tension along them, as is implied in Maxwell's statement.

[2] *Definition.*—The electric density at a given point on a surface is the limiting ratio of the quantity of electricity within a sphere whose centre is the given point to the area of the surface contained within the sphere, when its radius is diminished without limit.

by the movement of electricity, the phenomenon which we have called electric displacement is a movement of electricity in the same sense as the transference of a definite quantity of electricity through a wire is a movement of electricity, the only difference being that in the dielectric there is a force which we have called electric elasticity which acts against the electric displacement, and forces the electricity back when the electromotive force is removed; whereas in the conducting wire the electric elasticity is continually giving way, so that a current of true conduction is set up, and the resistance depends, not on the total quantity of electricity displaced from its position of equilibrium, but on the quantity which crosses a section of the conductor in a given time.

"That in every case the motion of electricity is subject to the same condition as that of an incompressible fluid, namely, that at every instant as much must flow out of any closed space as flows into it.

"It follows from this that every electric current must form a closed circuit."

It is quite possible that the difficulty in the concept of electric displacement may not be obvious on a first reading of this statement of the theory, as it only arises when we attempt to form some kind of mental picture of the actions taking place in the circuit. This is so greatly facilitated by the results of the electron theory that I shall anticipate some points which will be more fully considered in the next chapter. In a dielectric, the electrons cannot move freely, but they must be capable of a flux, or displacement, from their initial positions, under the action of electric force, until the latter is counterbalanced by the force which Maxwell calls the electric elasticity of the medium. The flux of the electrons must therefore give rise to a strain, and we shall see in the next chapter that this is not primarily a strain in the matter of the dielectric, but in the ether permeating it. The electric displacement, therefore, consists partly of this flux of electrons and partly of an ether flux or strain, and the two together produce a flux having the properties of the flow of an incompressible fluid; that is to say, a flux represented by a stream vector. The electron flux could be quite adequately represented as a polarisation of the particles or small elements of volume, since the electron displacement in each of these would electrify it positively on one side and negatively on the other. This polarisation is of the character considered in the older theories of polarity, and is not, by itself, a stream vector. The reader will therefore understand the difficulties experienced by von Helmholtz and other physicists who have attempted the further mathematical development of the consequences of the Faraday-Maxwell theory. I do not think there can be any doubt that it was this difficulty in conceiving the physical meaning of Maxwell's hypothesis of electric displacement that prevented the general acceptance of the

Faraday-Maxwell theory until its validity had been experimentally demonstrated by Hertz's brilliant experimental work.

In ether containing electrons the displacement will be of the same character as in a material dielectric, but if there are no electrons present, it will reduce to the ether strain alone, and this ether flux can be shown to be itself a stream vector, in the absence of electrons. This is the case in which the difficulty in the physical interpretation of electric displacement is most immediately obvious. We shall see in the next chapter that the electric displacement in free ether may be represented by a small twist in every volume element of ether in the electric field; now the ether strain arising from a small twist in every element of volume included in the portion of a circuit consisting of ether free from electrons might, indeed, be regarded as a polarisation, but it is physically very different from the true electric polarisation due to displacements of electrons.

In conductors some of the electrons are freely movable, so that there will be a true electric current due to a circulation of electrons. The ether flux due to the motions of these electrons will be insensibly small, so that the electron circulation in itself will sensibly form a stream vector. In addition to this circulation, there will be a polarisation due to displacement of electrons not free to move from molecule to molecule, and this polarisation will be accompanied, as in the case of a dielectric, by a flux of ether strain, the two together constituting a flux having the stream property.

In the light of the electron theory the physical meaning of electric displacement becomes perfectly clear, as will appear more fully in the next chapter. As set forth by Maxwell, it formed an admirable mathematical concept, perfectly adapted for the development of electrical theory at its then existing stage, but he was compelled to wait for a further advance in knowledge for its translation into a definite physical concept.

The general theory of polarity is again employed in developing the relations between magnetisation, magnetic force, and magnetic induction. The meaning of magnetic polarisation is explained by Maxwell in the following terms :—

"In speaking of the state of the particles of a magnet as magnetic polarisation, we imply that each of the smallest parts into which a magnet may be divided has certain properties related to a definite direction through the particle, called the Axis of Magnetisation, and that the properties related to one end of this axis are opposite to the properties related to the other end. The properties which we attribute to the particle are of the same kind as those which we observe in the complete magnet, and in assuming that the particles possess these

properties, we only assert what we can prove by breaking the magnet up into small pieces, for each of these is found to be a magnet."

The forces between magnets at rest, like those between electric conductors at rest, must necessarily be capable of derivation from a magnetic potential, since Green has shown that this is the necessary and sufficient condition for the system to obey the law of conservation of energy, viz. that the sum of the potential and kinetic energies of the system should be a constant quantity, a law which simply expresses in a definite form the denial of the possibility of perpetual motion.[1]

The magnetic potential at any point in a magnetic field is measured by the work done against the magnetic forces in bringing a unit magnetic pole from an infinite distance to the given point, assuming the unit to be small enough for the field to remain sensibly unaffected by so doing.

The simplest possible type of polar element is a doublet consisting of a positive point-pole of strength m, and an equal negative one separated from it by a very small distance l.

Let P be any point at a distance r from the positive pole, and r' from the negative pole, of such a doublet. Then the magnetic potential Ω at P will be m/r due to the positive pole, and $-m/r'$ due to the negative pole, so that the potential Ω of the doublet will be

$$\Omega = \frac{m}{rr'}(r' - r)\,.$$

If l is very small compared with r, we may put $r' - r = l \cos \epsilon$, where ϵ is the angle between the axis of the magnet and the straight line drawn from the particle to the point P, which gives

$$\Omega = \frac{ml}{r^2} \cos \epsilon\,.$$

The intensity of magnetisation of a particle being, as defined in Chapter I., the ratio of its moment to its volume, the magnetisation I at any point of a magnet may be defined by its intensity and direction.

The magnetic force in any given direction, at any point of a magnetic field, is equal to the rate of change of the magnetic potential at that point due to a displacement in the given direction, and the resultant force H at any point is the rate of change of the potential in that direction in which the rate is greatest. Magnetic force is, therefore, derived from magnetic potential in exactly the same manner that electric force is derived from electric potential.

If we wish to determine for any point either the electric potential of a given distribution of electric charges, or the

[1] See Appendix E.

magnetic potential of a given distribution of magnetic doublets, we can do so, when the distributions are continuous, by a similar mathematical process of summation, or adding up, of the potentials due to the separate particles in either case. In one respect the electrical theory is more general than the magnetic one, owing to the fact that the positive and negative electric charges can be separated, and their potentials derived independently of each other, while the positive and negative magnetic particles are always bound together in doublets which have to be treated as one. It is for this reason that the expression for the magnetic potential of a particle at a given point contains the squared reciprocal of its distance from the point, while the expression for the corresponding electric potential contains only the reciprocal. The result of this difference has led to a distinction having been made in the mathematical development, but Professor Minchin has pointed out to me that this is an error. In the electrical theory the summation can be carried out by a direct process both inside and outside conducting matter, the potential being shown to be finite and continuous throughout all space, but vanishing at an infinite distance from electric charges. The electric force, representing the rate of change of the potential with displacement, in the direction in which this rate is greatest, is also found to be finite throughout all space, and it therefore follows that the force derived from the electric potential is the whole electric force arising from any distribution of electric charges, even within regions where the distribution is continuous. The magnetic potential is also finite and continuous throughout all space, and vanishes at an infinite distance from all magnetic matter; but when it is attempted to determine the magnetic force from the magnetic potential by a mathematical method similar to that employed in determining the electric force from the electric potential, one of the possible methods of procedure appears to be legitimate only for points outside magnetic matter, and to break down for points inside. If this procedure were applied to determine the force inside magnetic matter, one of the elements in the summation would become infinite owing to its containing the reciprocal of the distance between the point in question and the magnetic doublet situated at the same point. This distance being zero, its reciprocal will be infinite. Under these circumstances, the mathematical method employed ceases to be a legitimate one, that is to say, it is no longer one which would necessarily lead to true results: Professor Minchin shows, however, that a very simple modification makes the procedure for determining the magnetic force from the magnetic potential a legitimate one for any point, inside, as well as outside, magnetic

matter (see Appendix F). A definite result could be obtained, with the unmodified procedure, by considering the force acting on a unit magnetic pole placed in a cavity within the material. Although the consideration of such a cavity is now no longer necessary for the derivation of the magnetic force from the magnetic potential, it is of advantage in enabling us to obtain a definition, of a simpler and more elementary character, of the magnetic force at a point within a magnet. It follows from the definition, on page 10, of a unit magnetic pole, that the magnetic force at a point in space, free from magnetic matter, is the force which would be experienced by a unit north pole placed at the point. In order to apply this definition to a point situated within magnetic matter, it would be necessary to imagine the magnetic matter to be removed from a small space surrounding the point. The cavity so formed must be small compared with the dimensions of the magnetic mass in which it is supposed to be made, but very large compared with the dimensions of a magnetic doublet; that is to say, compared with molecular magnitudes. There will then be a distribution of magnetism on the walls of the cavity which will in general give rise to a finite force on the unit magnetic pole placed within it, although it would not contribute a finite term to the magnetic potential. The value of this force will depend on the form of the cavity, and on the magnetisation at the place where the cavity is formed. Suppose the cavity to be hollowed out in the form of a cylinder, with its axis parallel to the direction of magnetisation, in a portion of the magnet in which the magnetisation is uniform in amount and direction. If the diameter is very small compared with the length, the magnetic force at the centre will not be affected by the surface distribution on the ends of the cylinder, so that its value will simply be $\mathbf{H}$, the magnetic force derived from the potential. Lord Kelvin called this the polar definition of magnetic force. If, on the other hand, the length be made very small compared with the diameter, so that the cylinder becomes a thin disc like a pill-box, the magnetic force due to the matter removed will be $-4\pi\mathbf{I}$.

The total force $\mathbf{H}$, due to the complete body, on the point at the centre of the box may then be considered as compounded of the force due to the matter inside the box and the force due to the matter outside the box. But the force due to the matter inside is $-4\pi\mathbf{I}$, so that if $\mathbf{B}$ is the force due to the matter outside the box, we shall have

$$\mathbf{H} = \mathbf{B} - 4\pi\mathbf{I},$$

and therefore,

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I}.$$

The quantities **H**, **I**, and **B** are clearly vectors, and must therefore, in the general case, be added by the law for compounding vectors, viz. from the point O, fig. 7, draw the straight line OP representing **H** in magnitude and direction, and from P draw PQ representing 4π**I** in magnitude and direction. Then OQ will represent **B** in magnitude and direction. If we first draw OR representing 4π**I**, and from R draw a straight line representing **H**, it must evidently pass through Q, showing that the order of addition is immaterial, as in the case of ordinary numerical quantities. In most cases, however, which have to be dealt with in practice, **H** and **I** are, as we shall presently see, in the same direction, and then they can be added numerically, giving rise to a vector **B** in the same direction as the other two. Lord Kelvin called this the electromagnetic definition of magnetic

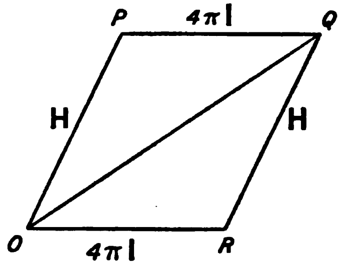


FIG. 7.

force. I shall follow the more convenient nomenclature of Maxwell, according to which **B** is defined as the magnetic induction and **H** as the magnetic force. 4π**I** is the local contribution to the magnetic induction.

Faraday's lines and tubes of magnetic force are, properly speaking, lines and tubes of induction, but the above investigation shows that, except in the interior of a magnetic substance, the two are identical. From the definition of unit tubes of magnetic force given in Chapter I., and which will also apply to the more general case of tubes of induction, it follows that the magnetic induction through any area is measured by the algebraic sum of the unit tubes of magnetic induction which traverse it. In a straight, uniformly magnetised bar the magnetic force due to the magnet itself is from the north pointing, or positive pole, towards the south pointing, or negative pole, both inside and outside the substance of the magnet; while the magnetic induction is from the positive to the negative pole, outside the magnet, and from the negative to the positive pole inside. The lines or

tubes of induction, therefore, form closed curves or tubes. If a closed curve is drawn enclosing a bundle of such tubes, then, remembering that the number of tubes passing through a surface is to be taken as determined by adding them algebraically, that is to say, by adding together all the tubes cutting the surface in the positive direction, and subtracting those cutting it in the negative direction, it will be seen that the number of tubes passing through any surface that can be drawn with the closed curve as boundary will be the same. The magnetic induction through a closed curve will therefore depend only on the closed curve, and not on the form of the surface of which it is the boundary. It should therefore be possible to determine the magnetic induction through a closed curve by a process depending only on the nature of the curve, and not involving the construction of a surface bounded by it. Maxwell does this by determining a vector **A** which is related to **B**, the magnetic induction, in such a way that the values of **A** summed up round the closed curve

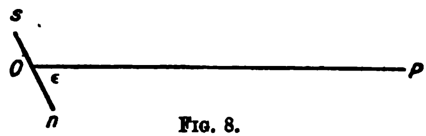


FIG. 8.

have a sum equal to the sum of the values of **B** summed up over a surface bounded by the curve. This quantity **A** he calls the vector potential of magnetic induction.

Let a magnetic doublet ns (fig. 8) of length l and poles m and $-m$ have its centre O joined to a point P, the line OP making an angle ϵ with ns. Then the vector potential, at the point P, due to the doublet at O, will be represented by a straight line drawn upwards through P, perpendicular to the plane of the paper, viz. to the plane containing OP and the axis of the doublet, of a length equal to $\frac{ml}{\mathrm{OP}^2}\sin\epsilon$. It will be observed that the numerical value of this vector is the ordinary scalar potential of the doublet.

When a body is magnetised by the action of magnetic force it is said to be magnetised by the induction of the magnetising force. All substances are found to give some indications of magnetic polarity when acted upon by a sufficiently strong magnetic force, but the magnetisation resulting from a given magnetising force varies greatly in different substances. If **I** is the induced magnetisation at a point where the magnetic force is **H**, then the ratio κ of **I** to **H** is called by Maxwell the coefficient

of induced magnetisation, and is now usually known as the *magnetic susceptibility*, so that $\mathbf{I} = \kappa\mathbf{H}$.

If the medium is isotropic, or if the direction of the applied magnetising force $\mathbf{H}$ coincides with an axis of greatest or least magnetic susceptibility, $\mathbf{I}$ and $\mathbf{B}$ will act in the same direction as $\mathbf{H}$, or in the opposite direction, as the case may be, and κ is then simply a number expressing the value of the numerical ratio I/H. In the ordinary cases of the magnetisation of iron occurring in practical work one of these conditions will always be fulfilled. In order that the above ratio may afford a true measure of the susceptibility for a given value of H, the material must be free from any magnetism except that which is induced by the force H. That is to say, there must be no residual magnetisation due to previously acting magnetic forces.

According to Ewing, the maximum value of κ for pure soft iron is about 245, and it is positive, so that the magnetisation is in the same direction as the magnetising force. Bodies for which this is the case are called paramagnetic, or simply magnetic, substances. In no other substance does κ approach the values attained in the case of iron, and for most paramagnetic substances, with the exception of nickel and cobalt, it does not exceed 1/100,000. Ewing gives 23·5 as the maximum value for nickel, and 13·8 for cobalt. Substances for which κ is negative are called diamagnetic substances, the most highly diamagnetic substance known being bismuth, and for this the value of κ is only about $-1/400{,}000$.

In crystals, bodies in a state of strain, and organic substances, the direction of magnetisation does not necessarily coincide with that of the magnetising force, and Maxwell shows that in the most general case six independent numbers are required to specify the complete relation between them. Phenomena of this kind were studied by Faraday under the name of "magne-crystallic action." When a crystal is freely suspended in a field of magnetic force it will tend to set itself with the axis of greatest paramagnetic induction, or of least diamagnetic induction, parallel to the lines of force at the place occupied by it.

Another coefficient of great importance is the ratio $\mathbf{B}/\mathbf{H}$, which Maxwell calls the magnetic inductive capacity of a substance, but which is now usually known as the *magnetic permeability*, the name given to it by Lord Kelvin. If we substitute κH for I in the equation

$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I},$$

we obtain

$$\mathbf{B} = (1 + 4\pi\kappa)\mathbf{H} = \mu\mathbf{H},$$

when μ is the magnetic susceptibility.

The similarity in the relations of magnetic force to free magnetism, and of electric force to free electricity, leads to the result that, generally speaking, the vector

$$\frac{\mathsf{B}}{4\pi}=\mathsf{I}+\frac{\mathsf{H}}{4\pi}$$

may be considered as the analogue, in the Faraday-Maxwell magnetic theory, of the electric displacement in the electric theory, and magnetic induction as the analogue of electrostatic induction, the magnetic permeability of a magnetic conductor corresponding to the specific conductivity of an electric conductor.

This analogy will be further considered in Chapter X., and its limitations indicated. Its basis is obviously the similarity between magnetic and dielectric polarisation, and I will now make use of it in order to deduce a relation between the dielectric constant and the density of a medium, which is of importance in connection with the refraction of electromagnetic waves, and also in accounting for the phenomena of "residual discharge."

Suppose the medium to be ether containing n material molecules per unit of volume, each polarised to the electric moment m by a field of electric force. The field is made up of the external exciting field and that of the polarised molecules, and of the latter portion, part is derived from the polarised medium as a whole, and part is local, depending only on the polarisation of the contiguous molecules. Let E′ be the local part, E the remainder of the total electric force, I′ the intensity of the polarisation, ρ the density of the medium, k a constant independent of the density, and p a parameter. Then we have

$$m=k(\mathrm{E}+\mathrm{E}') \text{ and } \mathrm{I}'=nm\,.$$

Moreover, E′ depends on nm, and therefore, since no limitations have been made regarding p, we may write $\mathrm{E}'=p.nm$.

Hence

$$\mathrm{I}'=kn(\mathrm{E}+p\mathrm{I}')\,.$$

Now, from the analogy considered above, if D is the numerical value of the electric displacement,

$$4\pi\mathrm{I}'+\mathrm{E}=4\pi\mathrm{D}=\mathrm{KE}$$

in electrostatic units, and therefore

$$\frac{\mathrm{K}-1}{\mathrm{K}-1+\dfrac{4\pi}{p}}=pkn\,.$$

In the case of fluid media it can be shown[1] that p is approximately equal to $\frac{4}{3}\pi$, and probably does not differ greatly from this value in the case of solid dielectrics, so that, approximately, $pkn = 4\pi kn/3$. Now ρ is proportional to n, and therefore we may write $pkn = A\rho$, approximately, where A is a constant. It follows that, for a given substance, the expression

$$\frac{K-1}{(K+2)\rho}$$

is approximately constant, a solution arrived at by Lorentz.

Again, the relation between K, E, and I′ may be written

$$K = \frac{4\pi I'}{E} + 1 = \frac{4\pi}{E}nm + 1\,,$$

which shows that in a given electric field the dielectric constant is increased, and consequently the electric elasticity of the medium diminished, by increasing the number of molecules, and therefore the number of electrons, in unit volume of the dielectric.

When a condenser, such as a Leyden jar, receives an electric charge, the strain arising from the polarisation may, if the electric force is sufficiently great, rupture the dielectric, and allow the condenser to discharge. The continuity of a fluid dielectric will then be restored, but the rupture of a solid one will be permanent. When the electric force ceases to act, without rupture taking place, the state of strain will disappear instantaneously in the case of fluids, but in solids the disappearance of the strain will be gradual, owing to the secondary elastic strains called into action, so that if the condenser is tested shortly after an instantaneous discharge, a small residual charge will be found. If gutta-percha is used as the dielectric, the time taken to return to the unstrained condition is considerable, and a series of residual charges, gradually decreasing in amount, may be obtained. In mica the effect is hardly perceptible, in glass it is much more strongly marked than in mica, and very much less than in gutta-percha. A similar time-lag of the polarisation behind the electric force occurs on the application of the latter.

Maxwell's theory of the residual charge attributed it to heterogeneity in the structure of the dielectric, which he assumed to consist of parts having varying conducting and dielectric properties, subject, however, to simple linear laws. This hypothesis leads to an exponential expression for the rate of recovery of the residual charge. No such formula, however, appears to

[1] See Appendix G.

fit the experimental results, and the subject has been recently investigated by the aid of refined methods by Trouton and Russ (*Phil. Mag.*, May 1907). These authors find that the recovery can be well represented by the logarithm of a quantity proportional to the time of recovery, a formula of the same type as that shown by A. O. Rankine (*Phil. Mag.*, 1908, vol. viii.), and others, to be applicable to the recovery from overstrain in elastic bodies, which is in accordance with the explanation given above.

It will be clear that the electric displacement through a surface whose boundary is any closed curve will, like the magnetic induction, remain unchanged if the surface is replaced by any other surface having the same curve for boundary, for the reasoning employed in the case of the magnetic induction depended only on its being a stream vector, and therefore applies to any quantity having that property.

The solid, or preferably *conical*, angle subtended at any given point by a closed curve is defined as being measured by the area of a spherical surface of unit radius described with that point as centre, the outline of which is traced out by a straight line (called a radius vector) passing through the given point and a moving point travelling completely round the closed curve. This area is reckoned positive or negative according as it lies on the left hand or on the right hand of an observer placed with an eye at the given point and always facing in the direction of motion of the point of intersection of the radius vector with the sphere.

We can now prove the statement made on page 120 that the introduction of a charge e within any finite closed surface gives rise to an outward displacement equal in amount to the charge. As is fig. 6, suppose the charge e to be uniformly distributed over a small sphere within, and concentric with, a sphere, which we shall now assume to have the radius unity, and suppose this sphere to be enclosed within a finite closed surface, as shown in section in fig. 9. Let the radius vector OPP′ trace out a small area, of which PQ is a section, on the sphere, and a small area, of which P′Q′ is a section, on the enclosing surface subtending a conical angle ω at O. The electric displacement, and therefore the electrostatic induction, which is 4π times the displacement, will be everywhere normal, that is to say, perpendicular, to the spherical surface, and therefore the cone of which P′POQQ′ is a section will be a tube of induction, so that the displacement, and therefore also the induction, through the area PQ, will be equal to that through the area P′Q′. This will hold good for every such elementary cone, and therefore for every corresponding pair of elements, of the enclosing surface and the sphere respectively. Therefore, the displacement through the outer closed surface is

equal to that through the spherical surface; that is to say, to e, which proves the statement.

Since the induction is 4π times the electric displacement, it follows that the total outward induction through a finite closed

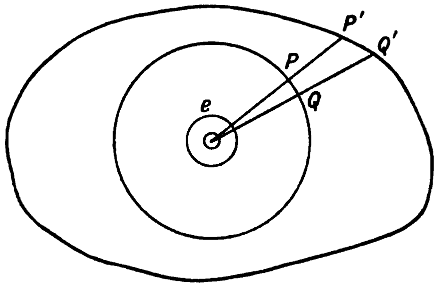


FIG. 9.

surface, due to an electric charge e within it, is $4\pi e$. It should also be noted that the induction through an area of the enclosing surface subtending a conical angle ω at O will be ωe, and therefore ω will represent the induction through an area subtending that angle at the point O, due to a unit charge at O. As this depends

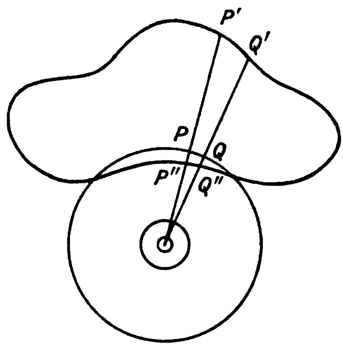


FIG. 10.

only on the charge e and the conical angle ω, the induction will be the same for any area subtending the same conical angle ω at O.

If the charge is concentrated round a point O outside the surface, as shown in fig. 10, each element PQ of the sphere will either give rise to no corresponding element of the closed surface, or to two elements P′Q″ and P′Q″, through which the inductions are equal in amount, but opposite in direction, one being outwards,

and the other inwards. The total induction outwards through the surface is therefore zero.

If the charge within the closed surface were $-e$, then there would be a total outward induction $-4\pi e$; that is to say, a total inward induction $4\pi e$.

The total induction though a closed surface described so as to include a number of electric charges will therefore be equal to the algebraic sum of the charges within it, and unaffected by those outside, and its direction will be outward if this sum is positive, and inward if it is negative.

There is one essential distinction between electrostatic induction and the corresponding electric displacement, on the one hand, and magnetic induction, on the other. The tubes of electrostatic induction do not form closed circuits, but terminate in a source, or positive electric charge, in one direction, and in a sink, or negative electric charge, in the other, while the tubes of magnetic induction form closed circuits, without either sources or sinks.

Every tube of magnetic induction intersecting a closed surface must therefore intersect it an even number of times, the induction through the elements of the surface successively outlined by the tube being directed inwards and outwards alternately. As the inductions through all elements due to the same tube are numerically equal, it follows that the induction through the surface due to any one tube is zero. It must therefore be zero for any number of tubes, so that the total magnetic induction through any closed surface is zero.

The reason for this difference is to be found in the fact that the positive and negative poles of a magnetic doublet cannot be separated, while those of an electric doublet can be separated to any desired distance, always remaining connected by a tube of electrostatic induction. Whenever, therefore, we consider the action of an isolated electric charge, we make the implicit assumption that the corresponding opposite charge is removed to an infinite distance; that is to say, to such a distance that it can have no effect on the problem under consideration. This explains the necessity of limiting the closed surfaces considered in electrostatic induction problems to finite surfaces. When we are dealing with the flux of induction through a finite closed surface, due to a charge within it or without it, we exclude the case of a tube of induction terminating within such a surface, or so near to it as to give rise to induction through it, for the charge at the further extremity would in that case not have been removed to a sufficient distance to eliminate its influence on the conditions of the problem. If the closed surface extended to infinity, we could no longer ensure the exclusion of the action of the remote end of the tube. No

such restriction is required in the case of the magnetic problem, owing to the tubes of magnetic induction necessarily forming closed circuits.

We have seen that the electric induction through a closed surface, due to a charge e within it, is $4\pi e$, which is simply the result of summing up, over the whole surface, the normal component of the electric force per unit of area. If the same thing is done for the magnetic force in the case of a closed surface containing a quantity m of free magnetism,[1] it follows from the similarity in the relations of H to m, and E to e respectively, that the sum so obtained must be $4\pi m$. Therefore, in order that the total flux of magnetic induction through the surface should be zero, since

$$\mathsf{B} = \mathsf{H} + 4\pi \mathsf{I},$$

the result of summing I all over the surface must be $-m$, which will give, as required, for the summation of B over the whole surface

$$4\pi m - 4\pi m = 0.$$

A mathematical proof of this, due to Professor Minchin, is given in Appendix F. The proof offered by Maxwell (treatise, § 402) is unsatisfactory in that it omits to take account of the magnetised particles lying on the surface.

If a thin shell of magnetic matter is magnetised in a direction everywhere normal to its surface, the product of the intensity of magnetisation into the thickness of the shell at any point is called the strength of the shell at that point. A magnetic shell in which the strength is the same at all points is called a Simple Magnetic Shell, and when the term magnetic shell is used without qualification it is to be taken to mean a simple shell. A magnetic shell in which the strength varies from point to point may be supposed to be built up by the superposition of a number of simple shells, and is called a Complex Magnetic Shell.

Let Φ be the strength of a simple magnetic shell, S′ the area of a small element of its surface at a point Q, and let r be the vector drawn from Q to any fixed point P, making an angle ϵ with the normal at Q on the positive side of the shell.

If l is the thickness of the shell, and m the strength of the positive pole of the element S′, its magnetic moment will be ml

[1] The quantity which has been called the strength of a magnetic pole may also be called a quantity of *magnetism*, provided that no properties are attributed to *magnetism* except those observed in the poles of magnets. The excess of north pole, over south pole, magnetism contained within a given space may then be called the free magnetism contained within the space.

and its volume will be lS', so that the intensity of magnetisation I at the point Q will be given numerically by the equation

$$I = \frac{\text{magnetic moment}}{\text{volume}} = \frac{ml}{lS'} = \frac{m}{S'}.$$

Hence

$$Il = \Phi = \frac{ml}{S'},$$

and therefore

$$ml = \Phi S'.$$

Thus, if Ω' is the magnetic potential of the element at the point P,

$$\Omega' = \frac{ml}{r^2} \cos \epsilon = \Phi \frac{S'}{r^2} \cos \epsilon.$$

But if ω' is the conical angle subtended by S' at the point P,

$$r^2 \omega' = S' \cos \epsilon;$$

so that

$$\Omega' = \Phi \omega',$$

and therefore, in the case of a simple shell,

$$\Omega = \Phi \omega,$$

where Ω is the magnetic potential of the whole shell at the point P, and ω is the angle subtended by the shell at that point.

That is to say, the potential due to a magnetic shell at any point is equal to the product of its strength into the conical angle subtended by its edge at the given point. It should be noted that ω numerically represents the magnetic induction due to a unit magnetic pole placed at P, a result which follows from the corresponding expression in terms of the electrostatic induction derived on page 134. This very important theorem, due to Gauss, shows that the magnetic potential of such a shell depends only on its strength, and on the curve forming its boundary, and not at all on the shape of the surface; and we shall see presently that this result makes it possible to reduce the determination of the mutual action of electric circuits to the simpler problem of determining the mutual action of magnetic shells.

By the definition of potential, Ω is the work done against the magnetic forces, due to the presence of the shell, in bringing a unit magnetic pole from an infinite distance to the point P. It follows therefore that the work done against these forces in bringing a magnetic pole m from an infinite distance to the point P is

$$m\Omega = m\Phi\omega,$$

which is therefore the potential energy at P due to the mutual action of the pole and shell.

If a magnetic pole m starts from a point on the negative surface of a magnetic shell and travels along any path in space so as to come round the edge to a point close to where it started from, but on the positive side of the shell, the conical angle will vary continuously, and will increase by 4π during the process. Suppose, for example, that the shell has a circular edge. The moving pole will start from a point within the circle, and on the negative side of its plane, and the angle subtended by the circle at the point will be represented by a hemisphere of the unit circle, its value being -2π. When the moving pole reaches a point in the plane of the circle, but outside it, the angle will have increased from -2π to 0, and as it returns to a point close to its starting point, but on the positive side of the plane, it will further increase from 0 to 2π. The total increase will therefore be 4π. The work done by the moving pole will therefore be $4\pi\Phi m$, and the potential at any point on the positive side of the shell will exceed that at a neighbouring point on the negative side by $4\pi\Phi$.

If a magnetic shell forms a closed surface, it will be seen that the potential outside the shell is everywhere zero, and in the interior is everywhere $4\pi\Phi$, being positive when the positive side is inward. Such a shell therefore gives rise to no magnetic force either at external or internal points.

In the expression $m\Phi\omega$ obtained above, as the mutual potential energy of a magnetic pole m at a point P and a magnetic shell of strength Φ subtending a conical angle ω at P, $m\omega$ is, as we have seen, the flux of the magnetic induction through the boundary curve of the shell Φ, due to the pole m at P. If the pole m is replaced by another magnetic shell Φ', we can obtain an expression for the mutual potential energy of the two shells by a summation of the expression $m\omega$ extending over the whole of the shell Φ', and multiplying the result by Φ. The value of a quantity M′, obtained in this manner, represents the product of Φ into the magnetic induction through its boundary due to the shell Φ'. It may also obviously be expressed as the product of Φ' into the magnetic induction through the boundary of Φ' due to the shell Φ. M′ must therefore be proportional to the product $\Phi\Phi'$ multiplied by a quantity obtained as a summation over surfaces bounded by the edges of the two shells. But we may, as we have seen, replace the summation representing the magnetic induction through either shell by the summation of the quantity which we have called the vector potential of the magnetic induction. M′ will then be proportional to the product of a summation over the two closed curves forming the boundaries of the shells, into their strengths. The result M of

this double summation is called the *mutual potential* of the two closed curves,[1] and is the fundamental quantity in the Faraday-Maxwell theory of the interaction of electric currents.

A mathematical evaluation shows that

$$M' = -\Phi\Phi' M,$$

so that the quantity M is the mutual potential energy, with its sign reversed, of two magnetic shells, each of unit strength, and having the two given curves as boundaries.

The negative sign is due to magnetic, like electric potential, being so defined that the resultant force in any direction is measured by the *decrease* of the potential in that direction. This makes it correspond in sign with potential energy, which always decreases when the bodies are moved in the direction of the forces acting on them.

It was first shown by Ampère, and confirmed by more accurate experiments by Weber and others, that the magnetic action of a small plane circuit, at distances which are great relatively to its dimensions, is the same as that of a magnet whose axis is normal to the plane of the circuit, and whose magnetic moment is equal to the area of the circuit multiplied by the strength of the current.

If the current C in such a circuit be replaced by a magnetic shell of strength C having the circuit as its boundary, then the magnetic action of the shell at all distant points will be identical with that of the current. This may be extended, by a method due to Ampère, to the case of the action of any electric circuit, of any form and size, upon any point P not in the conductor itself.

Suppose a surface to be described, bounded by the circuit, and not passing through P, and suppose two series of lines to be drawn upon it in such a manner as to divide it into elementary portions, of such small dimensions in comparison with their distances from P and with the radii of curvature of the surface, that they may be treated as small plane circuits in the sense before mentioned. Round each of these elements let a current of strength C be supposed to flow, the direction of circulation being the same in all the elements as it is in the original circuit. Then it will be seen that equal and opposite currents will flow along every dividing line between two adjacent elements, and

[1] *Mathemathical Note.*—If ds and ds' are elements of the two curves, ϵ the angle between them, and r their distance apart, then their mutual potential is

$$M = \iint \frac{\cos \epsilon}{r} ds ds',$$

the integration extending round both curves. The nature of the quantity M would have been more definitely indicated by terming it the *mutual potential coefficient* of the two circuits.

their effects must necessarily cancel each other. The only portions of the elementary circuits which are not cancelled in this manner are those which coincide with the original circuit. The total effect of the elementary circuits is therefore equivalent to that of the original one.

Now, each of the small circuits may be replaced by a magnetic shell of strength C bounded by the circuit, and the magnetic effect of this elementary shell on P will be the same as that of the elementary circuit. The whole series of elementary shells will form a magnetic shell of strength C having the original circuit as its boundary, and its magnetic action on P will be equivalent to that of the circuit. The direction of the current in the circuit will be so related to the magnetisation of the shell that it would appear to flow from right to left in front of a man standing on the positive, or north pointing, side of the shell.

There will, however, be an important difference between the magnetic potential due to the shell and that due to the circuit. The magnetic potential of the shell will have the value ωC at all external points, and within the substance of the shell it will vary continuously,[1] and have a single determinate value at every point.

In the case of the electric circuit, on the other hand, a moving magnetic pole may be made to thread through the circuit any number of times in either direction, with an expenditure of work amounting to $\pm 4\pi C$ each time, and therefore ωC, the value of the magnetic potential at every point not in the conductor itself, is a function having an infinite series of values differing by $4\pi C$. The derivative, or rate of variation, of ωC in any given direction has, however, a single determinate value at every point. That is to say, the magnetic force in any direction, due to an electric circuit, has a single determinate value at every point.

To determine the potential energy M′ of a magnetic shell of strength Φ placed in any magnetic field, we may consider the field as due to a number of magnetic poles $m_1, m_2, m_3, \ldots$ etc., placed at points at which the shell subtends solid angles $\omega_1, \omega_2, \omega_3, \ldots$ etc. Then we shall have

$$M' = \Phi(m_1\omega_1 + m_2\omega_2 + m_3\omega_3 + \ldots\ldots) = \Phi\Sigma m\omega, \text{ say.}$$

$\Sigma m\omega$ will then numerically represent the total magnetic induction

[1] At any point within the substance of a magnetic shell of strength ϕ we may suppose the shell to be divided into two parts whose strengths are ϕ_1 and ϕ_2 where $\phi_1 + \phi_2 = \phi$, and such that the point is on the positive side of ϕ_1, and on the negative side of ϕ_2. The potential at this point will then be

$$\omega(\phi_1 + \phi_2) - 4\pi\phi_2.$$

On the negative side of the shell the potential will be $\phi(\omega - 4\pi)$, so that the variation is continuous and the value singly determined.

through the boundary of the shell, due to the external magnetic field.

Let N be the algebraical sum of the unit tubes of magnetic induction which pass through the shell from the negative to the positive side. Then, since the shell does not belong to the system to which the induction is due, the magnetic force will be equal in both magnitude and direction to the magnetic induction, so that the induction is equal to the rate of greatest decrease of the potential, and therefore

$$M' = -\Phi N.$$

Let δr represent any displacement of the shell, and let F_r be the force acting on the shell in the direction of the displacement, and therefore tending to aid the displacement. Then, by the principle of the conservation of energy, the work $F_r \delta r$ done by the force during the displacement, added to the change $\delta M'$ in the potential energy, must be zero, viz.,

$$F_r \delta r + \delta M' = 0,$$

which gives

$$F_r = \Phi \frac{\delta N}{\delta r}.$$

The force corresponding to any given displacement of the shell will therefore assist or resist that displacement according to whether the displacement increases or decreases N, the number of tubes of magnetic induction which pass through the shell.

This must also hold for the equivalent electric circuit, so that any displacement of the circuit will be assisted or resisted according to whether it increases or decreases the number of tubes of magnetic induction which pass through the circuit in the positive direction.

Bearing in mind that the positive direction of a line of magnetic induction is the direction in which the north pointing pole of a magnet tends to move along the line, a line of induction will pass through the circuit in the positive direction when the direction of the line of induction is related to the direction of flow of positive electricity in the circuit, as is the longitudinal to the rotational motion of a right-handed screw.

If a portion of the circuit is flexible, so that it may be displaced independently of the rest, the correspondence of the circuit with a magnetic shell may still be maintained by making the edge of the shell capable of similar displacement, by cutting up its surface into a sufficient number of portions united by flexible joints. We may therefore conclude that any kind of change in form or position of the circuit which will increase the number of

tubes of magnetic induction passing through it will be aided by the electromagnetic force acting on the circuit. Every portion of the circuit is therefore acted on by a force urging it across the tubes of magnetic induction in such a manner as to include a greater number of these tubes within the circuit, and the work done by the force during this displacement is measured by the product of the number of tubes so added, into the strength of the current.

It follows from this that if the circuit be moved in any way so that, after assuming various forms and positions, it returns to its original place and shape, the strength of the current remaining unchanged during the motion, the whole amount of work done by the electromagnetic forces will be zero. It is therefore impossible to maintain by electromagnetic forces a continuous rotation of any part of a linear circuit carrying a constant current, against frictional and other resistance.

Continuous rotation may, however, be produced, provided that at some part of the course of the electric current it passes from one conductor to another which slides over it. In that case the circuit can no longer be considered a single linear circuit carrying a constant current, but must be regarded as a system of two or more circuits carrying variable currents such that those for which N is increasing have currents flowing in the positive direction, while those for which N is decreasing have currents in the negative direction.

If two circuits carrying constant electric currents C_1 and C_2 are supposed to be replaced by equivalent magnetic shells, then the potential energy due to their mutual action will be

$$-C_1C_2M,$$

where M is the quantity which we have defined as the mutual potential of the boundaries of the two shells, viz., in the present case, of the two circuits carrying the currents C_1 and C_2.

The force F_r which aids any displacement δr, will therefore be obtained from the equation

$$F_r = C_1C_2\frac{\delta M}{\delta r},$$

and this result contains the whole theory of the mutual action of electric currents. The quantity M may be measured experimentally, instead of being determined by mathematical calculation, and it is then usually known as the Coefficient of Mutual Induction of the two circuits.

Maxwell observes:—

"The mechanical force which urges a conductor carrying a current across the lines of magnetic force acts, not on the electric current, but on the conductor which carries it. If the conductor be a rotating disc

or a fluid it will move in obedience to this force, and this motion may or may not be accompanied with a change of position of the electric current which it carries. But if the current itself be free to choose any path through a fixed solid conductor, or a network of wires, then, when a constant magnetic force is made to act on the system, the path of the current through the conductors *is not permanently altered*, but after certain transient phenomena, called induction currents, have subsided, *the distribution of the current will be found to be the same as if no magnetic force were in action.* The only force which acts on electric currents is electromotive force."

The existence of the Hall effect, considered in Chapters XI. and XIV., shows that the statement in italics is incorrect. We shall see moreover that, according to the electron theory, this "mechanical force" arises from the electric forces between the electrons, and is therefore essentially an electric force.

All the phenomena of induction currents, as enumerated in Chapter I., may be summed up in the single statement:—When the number of tubes of magnetic induction which pass through the secondary circuit in the positive direction is altered, an electromotive force [1] acts round the circuit, which is measured by the rate of decrease of the magnetic induction through the circuit.

When the primary circuit is replaced by a magnetic shell, it must be borne in mind that the austral aspect of the shell corresponds to the positive aspect of the circuit, viz., the aspect in which the current appears to the observer to circulate in the direction of motion of the hands of a watch.

Let A and B be two circuits initially in the positions A_1 and B_1, and let them be moved simultaneously into new positions A_2 and B_2, with or without changes in their sizes and shapes, and let C_1 and C_2 be the initial values of the current in A in the initial and final position respectively. It has been shown, by means of suitably devised experiments, that in such a case, which represents the most general possible change in the relation of the primary to the secondary circuit, the total induction current in B during the simultaneous changes depends only on the initial state A_1, B_1, C_1, and the final state A_2, B_2, C_2, and not at all on the nature of the intermediate states through which the system may pass.

When there is no motion, that is to say, when $A_1 = A_2$, and $B_1 = B_2$, experiments show that the induced current is proportional

[1] Professor Minchin points out that we have here a very good illustration of the misleading character of the term electromotive force, which, as explained in Chapter I., is not a *force*, but a quantity of work per unit charge. The physical fact which this language is used to describe is simply that the induced current in a closed circuit of resistance R in a magnetic field is $\frac{dN}{dt}/R$. It would be meaningless to speak of a *force* acting round a circuit.

to the primary current. The current C must therefore enter as a simple factor into the expression for the total induced current, the other factor being a function of the forms and positions of the two circuits. It is found, moreover, that the value of this function depends only on the relative, and not on the absolute, positions of A and B, so that it must be capable of being expressed as a function of the distances of the different elements of which the circuits are composed, and of the angles which their elements make with each other. It is, in fact, the quantity which we have called the mutual potential of the two circuits, and which we have denoted by M.

Then, if R is the resistance of the secondary circuit, and M_1 and M_2 are the values of M in the initial and final position respectively, the total induction current in B will be

$$\frac{M_1C_1 - M_2C_2}{R}.$$

The total induced current in a given circuit depends only on the change taking place in the quantity MC, which change may arise, either from variation in the primary current C, or from any motion, of either the primary or secondary circuit, which alters M.

Maxwell points out that the conception of such a quantity, on the changes of which, and not on its absolute magnitude, the induction current depends, occurred to Faraday at an early stage of his researches. Faraday had had no mathematical training, though he appears to have possessed the most extraordinary intuitive perception of mathematical relations, and, without any aid from mathematical procedure, he was led to recognise the existence of something which is now known to be a mathematical quantity, and which, as Maxwell observes, may even be called *the* fundamental quantity in electromagnetic theory. Arriving, however, as he did, at this conception, by purely experimental methods, he ascribed to it a physical existence, and supposed the matter composing the secondary circuit, when in an electromagnetic field, to become modified into a peculiar condition, which he called the *Electrotonic State.* Maxwell was the first to recognise, "in the refined mathematical idea of the potential of two circuits, Faraday's bold hypothesis of an electrotonic state," which has always presented a great difficulty to those who have studied his *Experimental Researches.* Maxwell observes that "the scientific value of Faraday's conception of an electrotonic state consists in its directing the mind to lay hold of a certain quantity, on the changes of which the actual phenomena depend"; but, as he further points out, "without a much greater degree of development than Faraday gave it, this conception does not easily lend

itself to the explanation of the phenomena," and the method of representing the electromagnetic field by means of its lines of magnetic force was, in his hands, far more powerful and fertile. The number of unit magnetic tubes which pass through the circuit at any instant is mathematically equivalent to Faraday's earlier conception of an electrotonic state, and it is represented by the quantity MC.

It was observed by Fleeming Jenkin that, if the current from a single voltaic cell is made to pass through a coil of insulated wire wound upon a soft iron core, forming an electromagnet, and the circuit is broken between the ends of two wires held in the hand, a smart shock is felt, but that no shock is felt on making contact, nor can any shock be obtained from the cell without the aid of the electromagnet.

Faraday showed that this, and other phenomena described in his *Experimental Researches*, are due to the same inductive action as that which he had found to be exerted by the current on neighbouring conductors, and this inductive action of a current upon the flow of electricity in the same circuit he called *self-induction.*

The coefficient of self-induction of a circuit can be calculated by means of the mathematical formula for the coefficient of mutual induction of two circuits, by making the two circuits coincide with each other. It can also be measured experimentally, and is usually denoted by the letter L.

Faraday observes that "the first thought that arises in the mind is that the electricity circulates with something like momentum or inertia in the wire." Indeed, when only a single wire is considered, the phenomena caused by the breaking of the current are exactly analogous to those arising from the sudden stoppage of the flow of water in a pipe, and which are due to the momentum of the moving mass of water. When, however, the phenomena are compared in their totality, a very striking distinction is observed. The effects of the fluid momentum depend only on the mass of the fluid flowing in the tube, on the length of the tube, and on its sectional area at different parts of its length. As long as the length of the tube remains unaltered, the phenomena are unaffected by the form into which it is bent, or by anything outside the tube.

This is not by any means the case for the wire carrying an electric current. If the circuit consists of a long wire doubled on itself, the effect is very small, but may be increased to some extent by separating the portions in which the current flows in opposite directions. A very much greater increase is observed when the circuit is bent into a helix, so that a number of turns of

wire in which the current flow is parallel, and in the same direction, are brought near together. The greatest increase of all is obtained when a soft iron core is placed within the coil, the effect of which, as we know, is to increase the number of tubes of magnetic induction which pass through the interior of the helix. Moreover, if two wires insulated from each other are coiled up together, then if the second wire does not form a closed circuit, the phenomena observed on breaking the first circuit, while traversed by an electric current, are unaffected by the presence of the second circuit; but if the second wire forms a closed circuit, then the effects of the self-induction in the first wire are diminished by the action of the induced current in the second wire. It follows therefore that if the results are due to momentum, the momentum cannot be that of the electricity in the wire. Then again, when an E.M.F. is applied to a circuit, the current does not at once attain its full value; and when the E.M.F. is suddenly removed, the current does not cease at once, but continues to flow until stopped by the resistance. Before it is stopped, it will have generated a certain amount of heat, the energy of which must have been derived from the kinetic energy of the current. The stoppage may be accelerated by allowing the current to do mechanical work, *e.g.* by moving magnets, the inductive effect of these motions causing the current to die away more rapidly than when the resistance of the circuit was the only cause tending to stop it.

It follows, therefore, that a system containing an electric current must be a seat of energy of some kind, and since an electric current can only be conceived as some kind of kinetic phenomenon, this energy must be kinetic; that is to say, the energy possessed by a substance in virtue of its motion. Since this energy cannot be entirely due to the motion of the electricity in the wire, and further experiments show that if any of the energy at all is due to this source, it must be an extremely small proportion, too small apparently to be detected experimentally, we are naturally led to look for some motion in the space surrounding the wire, whether occupied by a material dielectric or only by ether.

Maxwell is now in a position to develop his general scheme for reducing the phenomena of electric currents to a purely dynamical basis, starting from the expression of the electrodynamic potential of a system of currents, first discovered and employed by F. E. Neumann, and which is represented, in the case of only two circuits, by the quantity which we have called M.

In such a procedure the only possible course appeared to Maxwell to be to take the current flowing in an element of

volume, together with the co-ordinates which determine the position of the element, as the independent variables in a continuous mathematical analysis extended over the field of currents.

In the method employed by Neumann, and further developed by von Helmholtz, the function in question is considered simply as a potential from which the forces in the field are derived. It is then assumed that any variation in the potential, due to changes in the positions of the conductors, the currents remaining constant, gives rise only to pondero-motive forces tending to alter their positions; while a variation due to changes in the currents, whether the conductors be fixed or moving, gives rise only to electric forces tending to alter these currents. The validity of this assumption was experimentally demonstrated by Maxwell.[1] He further assumes that the potential function with its sign changed represents the kinetic energy of the ether associated with the system of currents. The induced electric forces are then taken to be the reversed kinetic reactions corresponding to the currents considered as generalised electric velocities, a perfectly valid proceeding, since an electric current represents the rate of change of a quantity of electricity relatively to the time, and these, together with the externally applied electric forces, together give rise to the currents in accordance with Ohm's law.

The pondero-motive forces acting on the conductors are similarly taken to be the reversed kinetic reactions corresponding to the changes in the positions of the conductors, which have to be equilibrated by equal and opposite applied forces in order to maintain the mechanical equilibrium of the system. Maxwell assumes with Ampère that the mechanical force of electrodynamic origin acting upon any element of a linear conductor carrying a current is at right angles to the element, which makes the problem a definite one for a system of currents flowing in closed circuits. For such currents Maxwell shows, by means of this assumption, that the potential can be represented as the kinetic energy of the medium connected with, and influencing, the material conductors, in a scheme in which the currents are treated as generalised velocities corresponding to electric co-ordinates which do not themselves appear in the mathematical expression of the function.

The system is then one to which the principle of least action can be applied, so that the general equations of the electromagnetic field can be obtained by the ordinary Lagrangian methods.

Larmor has shown that a general mathematical investigation

[1] We now know that this assumption can only be approximately correct.

on the basis of regarding the current elements as independent entities, but without the assumption made by Maxwell that the resultant electrodynamic force on each linear element is at right angles to the element, leads to a system of forces differing from the Ampère-Maxwell system, and one which leads to results which are in conflict with those obtained experimentally. Considering this in conjunction with the fact that the theory of moving electrons, which are certainly independent physical entities, leads, without any supplementary hypothesis or experimental principle, to the Ampère-Maxwell system with a mechanical force at right angles to each current element, Larmor concludes, and seems to be justified in the conclusion, that the specification of an electromagnetic system by means of independent current elements is physically insufficient. Maxwell's demonstration of the possibility of a mechanical theory of electric actions cannot therefore be considered as theoretically complete in itself, for the assumption that the resultant electrodynamic force on each linear current element is at right angles to the element was a purely empirical one. It was not derived from dynamical considerations, and its validity was based merely upon its leading to results in accordance with those experimentally observed. The demonstration is, however, rendered theoretically complete by the electron theory, in which the currents are regarded as convection currents due to the movements of discrete electrons, separated by intervening spaces containing only ether, completed and made circuital by means of the ether strains arising from the movements of the electrons, since, from this point of view, no empirical assumption is required, the actions between the moving electrons, and the accompanying variations in the ether strain, being derived from purely dynamical considerations.

In the course of his investigation Maxwell shows that the electromotive force in any circuit, arising from electromagnetic action alone, is equal to the rate of decrease of a quantity of the nature of a momentum, which he calls the electromagnetic momentum of the circuit. This quantity therefore represents the number of tubes of magnetic induction which cross the circuit, that is to say, it is the magnetic induction **B** through any surface bounded by the circuit. From this he derives the vector **A**, which has been defined as the vector potential of magnetic induction.

This vector represents in direction and magnitude what is called in mathematical language the time-integral or time-sum of the electric force which would be experienced by a particle carrying a unit positive charge placed at the point of its origin if the primary circuit were suddenly broken. He therefore calls

it the electrokinetic or electromagnetic momentum at the point. This time-sum of the electric force is obtained by dividing the time of action of the force into a number of small equal intervals, multiplying each elementary interval by the average value of the electric force during the interval, and taking the sum of all these products. In the case of a suddenly exerted mechanical force, as exemplified by a blow, this time-sum of the force is known as its total *Impulse*, and affords a measure of the effect of the blow. The electrokinetic momentum at any point due to any system of electric currents may be defined as the impulse of the electric force acting at the point when the currents are suddenly stopped. The quantity $\mathbf{A}$ may now be defined as the vector potential of the electric current, standing in the same relation to the electric current that the ordinary scalar potential stands to the matter of which it is the potential, and is obtained by a similar mathematical process which may be described as follows:—

Let a vector be drawn from any fixed point representing in direction and magnitude an element of the current, divided by the numerical value of its distance from the fixed point. When this is done for every element of the current, the resultant of all the vectors so obtained is the potential of the whole current, and the current being a vector quantity, its potential is also a vector.

By expressing the electromagnetic momentum of the secondary circuit as a vector summation of the vector potential round the circuit, and deriving its rate of change with the time, which is equal to the induced electromotive force round the circuit with its sign changed, a general expression is obtained for the latter, and from this, by means of some mathematical transformations, the electromotive force round the circuit is obtained in the form of a summation round the circuit of an expression depending on the components of the magnetic induction parallel to the three rectangular axes at a point on the circuit, the time rates of the co-ordinates, and the time rates of the components of the vector potential. Maxwell adds to these the time rates of an undetermined function Ψ, which disappears from the result when the summation is taken round a closed circuit, in order to provide for the representation of further phenomena which might be experimentally observed. This quantity Ψ is found, on the electron theory, to represent the electrostatic potential of the whole of the electrification and polarity of the field. When

this potential is got rid of by elimination, there remains only a scheme of stream vector relations which suffices to render the determination of the sequence of changes a perfectly definite problem, so long as the conductors and dielectrics are at rest in the ether, so that in this case Ψ is not an independent variable, and its introduction is, in fact, unnecessary. When the material media are moving through the ether, Maxwell's scheme requires modification.

Ampère's explanation of molecular magnetisation as due to electric currents circulating in each molecule brings magnetic phenomena without further consideration within the purview of this general scheme of electrodynamic theory evolved directly from fundamental dynamical principles, and from the theoretical point of view introduces a very considerable simplification. The magnetic force becomes the same inside a magnet as in the space outside, and is, like the magnetic induction, a stream vector, and the three vectors, the electromagnetic momentum, the magnetic force, and the electric current, have their mathematical relations very greatly simplified.[1]

In the electrodynamic theory, developed on the single basis of moving electrons, the reason for this simplification becomes very clear. In a current system devoid of magnetism the only magnetic quantity that occurs is the magnetic induction due to the currents, as the portion of this induction arising from contiguous molecules or elements of volume is always negligible in comparison with the induction as a whole. Magnetic force is therefore not a fundamental quantity in an electrodynamic theory developed on this basis, but is a concept introduced in passing from molecular dynamics to the mechanical theory of continuous masses, and represents an averaged etherial disturbance diminished by the portion which is due to purely local causes.

Maxwell points out that if we suppose our mathematical machinery so coarse that our line of summation cannot thread a molecular circuit, and that an immense number of magnetic molecules are contained in our element of volume, we must still retain the old magnetic theory with its distinction between the force

[1] *Mathematical Note.*—They are all three stream vectors, and each of the two latter is obtained as the space-variation of the preceding one, viz.,

$$\mathbf{H} = \nabla \mathbf{A}, \text{ and } 4\pi \mathbf{C} = \nabla \mathbf{H};$$

where ∇ is Hamilton's operator

$$\nabla = i\frac{d}{dx} + j\frac{d}{dy} + k\frac{d}{dz},$$

i, j, k representing, as usual, unit vectors in three directions at right angles to each other.

within magnetic matter and in the space outside; but if we suppose our machinery of a finer order and capable of investigating all that goes on in the interior of the molecules, we must give up the old theory of magnetism and adopt that of Ampère, which admits of no magnets except those which consist of electric currents.

The development of our mathematical machinery has, as a matter of fact, advanced so far since the date of Maxwell's treatise that we are at present able to investigate to a very considerable extent the processes going on in the interior of molecules. It has not, however, advanced to a sufficient extent to make it in all cases a convenient instrument for dealing with magnetic phenomena on the purely electrodynamic basis. We are obliged to a very great extent to use methods capable only of dealing with groups of molecules aggregated into elements of volume, and the results so obtained are the averaged results obtained by the summation of the effects due to the separate molecules.

Moreover, in very many cases it is these averaged results which are required for the solution of practical magnetic problems, so that even if our mathematical machinery were to attain a sufficient development to enable us to study the actions of individual molecules in every case, we should then have to average the results in order to obtain the data required. In such cases no advantage would be gained by the use of the more elaborate analysis, so that we should still prefer the short cut provided by the methods of the ordinary magnetic theory.

If this theory had been developed in the light of our present knowledge, there would have been in all probability little or no modification in the methods themselves, but in that case what is now called the magnetic force at a point in the interior of a magnet would certainly have been defined, as it should be, as the part of the force acting on a unit pole at the point, which is independent of the local polarity, and it would then be shown how the definite value of this local part can be determined. It would only introduce confusion if we were now to apply the term magnetic force to what has been so long known as magnetic induction.

When we considered the mutual energy of two magnetic shells, the action between them was treated as an experimentally demonstrated fact, without attempting to account for it. The energy of a magnetic system was therefore assumed to be potential energy, which energy is diminished when the parts of the system are displaced by the action of the magnetic forces. The coefficient of mutual induction of the two shells was therefore assumed to be negative, while the coefficient of mutual induction of two electric circuits was taken to be positive, since the energy of a system of electric currents is kinetic, and is therefore increased by displace-

ments arising from the actions between them. If, however, magnets are regarded as deriving their properties from electric currents circulating within their molecules, their energy must be kinetic, and the forces acting between them must therefore be such as to tend to move them so that if the currents in them were maintained constant, this kinetic energy would increase, and therefore the coefficient of mutual induction of two magnetic shells will become identical in sign as well as in magnitude with that of the corresponding electric circuits.

In considering, therefore, the total energy existing in an electromagnetic field, it will not be necessary to take account separately of the energy due to the presence of magnetic matter, as this will be included in the general expression for the electromagnetic, or kinetic, energy of the complete system of electric currents, including the molecular currents in the magnetic matter.

The energy of an electromagnetic field will therefore consist of two parts only, the electrostatic, or potential energy, and the electromagnetic, or kinetic energy. To determine the former, suppose the system of electrification to be divided into a number of parts, each of which is so small that the potential may be treated as constant throughout its extent, and let e_1, e_2, etc., be the electric charges in each of these parts, and V_1, V_2, etc., the corresponding potentials. If e_1, e_2, etc., are changed into ne_1, ne_2, etc., the potentials will become nV_1, nV_2, etc. Now let the charges ne_1, ne_2, etc., be increased by $e_1\delta n$, $e_2\delta n$, etc., where δn is a very small quantity, these charges being supposed to be brought from a distance at which the electrical action of the system is insensible. The work done in adding the charge $e_1\delta n$ to ne_1, whose potential is nV_1 before the addition, and $(n+\delta n)V_1$ afterwards, must lie between

$$nV_1e_1\delta n \text{ and } (n+\delta n)V_1e_1\delta n.$$

If δn is taken small enough, its square may be neglected, so that the work done will be

$$V_1e_1n\delta n.$$

The work done in increasing ne_1, ne_2, etc., to $(n+\delta n)e_1$, $(n+\delta n)e_2$, etc., will therefore be

$$(V_1e_1 + V_2e_2 + \text{ etc.}),\ n\delta n = n\delta n\Sigma Ve, \text{ say}.$$

If this process is repeated an indefinitely great number of times, each charge being indefinitely small, until the total charge becomes sensible, the work done will be

$$\Sigma Ve\int_{n_0}^{n_1} ndn = \tfrac{1}{2}\Sigma Ve(n_1{}^2 - n_0{}^2),$$

where dn is written for the indefinitely small value of δn, and n_0 and n_1 are the initial and final values of n, respectively. If we make $n_0 = 0$ and $n_1 = 1$, we find as the work W done in charging an unelectrified system, so that the charges are e_1, e_2, etc., and the potentials V_1, V_2, etc., in each element,

$$W = \tfrac{1}{2}\Sigma Ve,$$

the summation being extended to every place where there is electrification.

Maxwell then expresses this sum in terms of the components of electric displacement and those of electric force, and shows that the same result will be obtained by a summation throughout the whole of space, instead of confining it to regions of electrification. That is to say, the electrostatic energy of the whole field will be the same when it is regarded as residing in every part of the field where electric force and electric displacement occur, as when it is treated as being confined to places of free electrification. The latter process would be appropriate to the theory of action at a distance, the former one to the theory which seeks to explain these actions by the agency of an intervening medium.

When expressed in this way, it is found that the electrostatic energy in unit of volume is represented by half the product of the numerical values of the electric force and the electric displacement, multiplied by the cosine of the angle between these vectors. That is to say, if E and D are the numerical values of the vectors $\mathbf{E}$ and $\mathbf{D}$ and ϵ is the angle between them, the electrostatic energy per unit of volume is

$$\tfrac{1}{2}ED\cos\epsilon.$$

In order to obtain an expression for the electrokinetic energy of a system of electric currents, let C_1, C_2, etc., be the currents, and p_1, p_2, etc., the electrokinetic momenta of the corresponding circuits.

If the currents are now changed into nC_1, nC_2, etc., the momenta will become np_1, np_2, etc.

Now let n be increased by δn, then the work done in increasing the kinetic energy of the system will be

$$nC_1p_1\delta n + nC_2p_2\delta n + \text{etc.}$$

Therefore, if n is made to increase from 0 to 1, the system will be brought from rest into the state of motion ΣpC, and the whole work done in increasing the kinetic energy will be

$$(C_1p_1 + C_2p_2 + \text{etc.})\int_0^1 n\,dn.$$

But
$$\int_0^1 ndn = \tfrac{1}{2},$$
and the work done in producing the motion is equivalent to the kinetic energy, so that if this be denoted by T,
$$T = \tfrac{1}{2}(C_1p_1 + C_2p_2 + \text{etc.}) = \tfrac{1}{2}\Sigma Cp.$$

If L_1, L_2, etc., are the coefficients of self-induction of the circuits, and M_{12}, etc., are their coefficients of mutual induction, this equation may be written
$$T = \tfrac{1}{2}(L_1C_1^2 + L_2C_2^2 + \text{etc.} + M_{12}C_1C_2 + M_{21}C_2C_1 + \text{etc.}).$$
Or, since $M_{21} = M_{12}$,
$$T = \tfrac{1}{2}L_1C_1^2 + \tfrac{1}{2}L_2C_2^2 + \text{etc.} + M_{12}C_1C_2 + \text{etc.}$$

Taking the expression in the form
$$T = \tfrac{1}{2}\Sigma Cp,$$
Maxwell then expresses it in terms of the components of the vector potential and those of the electric currents, summed up round all the circuits of the system, and shows that the expression so obtained may be transformed into one in terms of the components of magnetic induction and of magnetic force, the summation of which is extended throughout the whole of space.

This shows that the electrokinetic energy of the whole field will be the same when it is regarded as residing in every part of the field where magnetic induction and magnetic force occur, as when it is treated as being confined to the circuits carrying electric currents.

Here again the latter method would be appropriate to the theory of action at a distance, which simply presents a scheme representing the relations between the actions without attempting to account for them further, while the former method is the one appropriate to our present theory, which aims at accounting for these actions by the agency of the medium known as the ether.

The electrokinetic energy per unit of volume is found in this way to be
$$\frac{1}{8\pi}BH\cos\epsilon$$
where ϵ is the angle between the magnetic force and the magnetic induction.

When there is no magnetisation this becomes
$$\frac{1}{8\pi}H^2.$$

Maxwell then proceeds to apply these results to the determina-

tion of the mechanical forces acting on each element of a body placed in an electromagnetic field, and then expresses the force system, or *forcive*, so arrived at, in terms of the stresses produced in the etherial medium. His method of analysis involves the implicit assumption that the polarisation current is insignificant in amount in comparison with the true, or conduction, current. This is usually admissible, except in the case of radiation, but this limitation was overlooked by Maxwell, with the result that he arrives at expressions, in paragraph 792, on the Energy and Stress of Radiation, which are only correct under limitations which will be considered in Chapter XII.

Subject to this limitation, if 2ϵ be the angle between **H** and **B**, the mechanical electrokinetic forces acting upon any part of a material system at rest would, except for an outstanding couple or *torque*, equal in amount to $\frac{1}{4\pi}\mathrm{BH}\sin 2\epsilon$, tending to turn every element of volume from **B** towards **H**, be statically equivalent to a traction over its boundary specified as follows:—

(1) A pressure equal in all directions, or hydrostatic pressure, $=\frac{1}{8\pi}\mathrm{H}^2$.

(2) A tension along the bisector of the angle between **B** and **H** $=\frac{1}{4\pi}\mathrm{BH}\cos^2\epsilon$.

(3) A pressure along the bisector of the supplementary angle $=\frac{1}{4\pi}\mathrm{BH}\sin^2\epsilon$.

When **B** and **H** are in the same direction, as is the case in isotropic media which are not permanently magnetised, the torque vanishes, so that the representation of the mechanical electrokinetic forces in the form of a stress is perfect, and the traction on any surface, due to the stress, is then a hydrostatic pressure $\frac{\mathrm{H}^2}{8\pi}$ together with a tension $\mu\frac{\mathrm{H}^2}{4\pi}$ along the lines of magnetic force.

The theory of a polarised dielectric leads in a similar manner to the result that in isotropic dielectrics the mechanical electrostatic forces are derivable from a stress consisting of a hydrostatic pressure $\frac{\mathrm{E}^2}{8\pi c^2}$ together with a tension $\frac{\mathrm{KE}^2}{4\pi c^2}$ along the lines of electric force **E**, to which, however, is to be added a pressure $\frac{(\mathrm{K}-1)\mathrm{E}^2}{8\pi c^2}$ acting over the surface of every conducting region.

These differ from those given by Maxwell, owing to the fact that he considers the function of a uniform dielectric to consist merely in transmitting the forces without adding anything to them. The theory now being considered, however, which affords a basis for a physical explanation of electric displacement, regards a material dielectric as being polarisable similarly to a magnetic substance, so that the medium itself will afford a contribution to the mechanical forces.

A very important theorem with regard to the energy of the electromagnetic field has been obtained by Poynting,[1] expressing the fact that any change in the energy, due to all sources, contained within any closed surface may be regarded as arising from a transference of energy across the surface in a direction perpendicular both to E and H. The direction of the flow of energy is such that the rotation of a right-handed screw from E to H would be accompanied by a translation in the direction of the flow. The amount of energy which crosses unit area perpendicular to the direction of flow in a second is

$$\frac{1}{4\pi}\mathrm{HE}\sin\theta$$

where θ is the angle between H and E.

The solution obtained by the mathematical analysis not being a unique one, does not show that the flow of energy at any point *must* be that given above, but it does show that correct results with regard to the changes in the distribution of energy in the field will be arrived at by treating the energy as flowing in accordance with these laws.

It is pointed out by Larmor that Poynting's theorem is strictly applicable only in cases where the configuration of the electrons is not changing, so that for its strict validity, the bodies within the region to which it is applied, which carry currents or electric charges, or which are polarised, must be at rest, and there must be no change in the state of electrification of any conductor in the region. It is not applicable when the energy of electric or magnetic polarisation is incompletely organised, and mixed up with other molecular energy of the material, as it would be if there were hysteresis (see Chapter X.) or permanent magnetism.

Maxwell's last great contribution to the development of electrical theory is his electromagnetic theory of light, the formulation of which was in all probability due to Faraday's discovery of the effect of a magnetic field in producing a rotational displacement, or twist of the plane of polarisation, of a plane polarised ray of light. He approaches the subject in his treatise by observing that the

[1] *Philosophical Transactions of the Royal Society*, 1884, part II., p. 343.

undulatory theory of light, in common with the electrical theory there developed, assumes the transmission of actions by means of a medium, and that it would be an unphilosophical proceeding to assume space to be filled with a new medium whenever newly discovered phenomena require explanation.

When light is emitted, a certain expenditure of energy by the luminous body takes place, and if the light is absorbed by another body, the latter becomes heated, showing that it has received energy from outside. During the interval of time after the light left the first body, and before it reached the second, it must have existed as energy in the intervening space. According to the undulatory theory of light, this energy is passed on, by the action of contiguous parts of the medium, from one portion to the next, until it reaches the illuminated body.

The luminiferous medium is therefore, during the passage of light through it, a receptacle of energy, which must be supposed to be partly potential and partly kinetic, and in this respect the properties which must be ascribed to it are also those which must be ascribed to the electromagnetic medium, which must also have the capacity of becoming a receptacle of these two forms of energy. In both cases, the potential energy being supposed due to the distortion of the elementary portions of the medium, the latter must be elastic; and the kinetic energy being supposed due to the vibratory motion of the medium, it must have a finite density.

Maxwell then proceeds, by expressing the total current, made up of the conduction current and the variation of the electric displacement, in terms of the vector potential of the current system and the electrostatic potential of the electrification, to obtain equations determining the conditions of propagation of an electromagnetic disturbance through a uniform medium supposed to be at rest, that is to say, to have no motion except that which may be involved in electromagnetic disturbances.

In a dielectric medium these equations assume a simple form which shows that an electromagnetic disturbance is transmitted with a definite velocity v, given by the equation

$$v = \frac{1}{\sqrt{K\mu}}.$$

If the medium is air or free ether, $\mu = 1$, and $K = \frac{1}{c^2}$, where c is the radiation constant of the ether, so that we have $v = c$, the velocity of light in air or free ether. Maxwell's argument from this I will give in his own words (treatise, § 788):

"In other media than air, the velocity v is inversely proportional to the square root of the product of the dielectric, and the magnetic,

inductive capacities. According to the undulatory theory, the velocity of light in different media is inversely proportional to their indices of refraction.

"There are no transparent media for which the magnetic capacity (the permeability) differs from that of air more than by a very small fraction. Hence the principal part of the difference between these media must depend on their dielectric capacity. According to our theory, therefore, the dielectric capacity of a transparent medium should be equal to the square of its index of refraction.

"But the value of the index of refraction is different for light of different kinds, being greater for light of more rapid vibrations. We must therefore select the index of refraction which corresponds to waves of the longest periods, because these are the only waves whose motion can be compared with the slow processes by which we determine the capacity of the dielectric."

Maxwell then makes the observation that the only dielectric of which the capacity had then been determined with sufficient accuracy was paraffin, for which Gibson and Barclay had found that $K = 1{\cdot}975$. From Dr Gladstone's observations of the index of refraction for several wave-lengths, he deduces the index for very long waves as 1·422, while the square root of $K = 1{\cdot}405$.

Maxwell observes:—

"The difference between these numbers is greater than can be accounted for by errors of observation, and shows that our theories of the structure of bodies must be much improved before we can deduce their optical from their electrical properties. At the same time, I think that the agreement of the numbers is such that if no greater discrepancy were found between the numbers derived from the optical and the electrical properties of a considerable number of substances, we should be warranted in concluding that the square root of K, though it may not be the complete expression for the index of refraction, is at least the most important term in it."

A great number of such comparisons have been made since this was written by Maxwell, and many of the discrepancies were found to be very much larger than in the case of paraffin, some being so large as to fail to indicate any connection between the two constants. These results were for a long time considered to tell strongly against the validity of Maxwell's electromagnetic theory of light. When, however, it is borne in mind that the indices of refraction can only be observed in the case of the extremely short luminous waves; that the indices for very long waves have to be estimated from these by a mathematical formula which depends for its validity on the absence of anomalous dispersion in the waves of greater length than the visible ones; that all the earlier determinations of K had to be made in steady electric fields; and that Maxwell's equations took no account of the phenomenon of dispersion—the marvel is, as Sir J. J. Thomson has pointed out, not that there should be substances for

which the relation $K = n^2$ (where n is the index of refraction) does not hold, but that there should be any for which it does, and that, to give the theory a fair trial, the specific inductive capacity should be measured for electrical waves of the same length as the luminous waves employed to determine the refractive index.

Nevertheless, the relation was found to be approximately true for gases, and also for a considerable number of denser bodies as well, and some determinations made by Rubens and Arons in 1890, in which K was determined, not in a steady electric field, but by means of the shortest electric waves that could be produced, gave results very closely agreeing with Maxwell's formula, viz.,

	n	$\sqrt{K}$
Castor oil	2·05	2·16
Olive oil	1·71	1·75
Xylol	1·50	1·53
Petroleum	1·40	1·44

It is shown on page 160 that the velocity of displacement of the rotationally elastic ether may be taken as representing magnetic induction or magnetic force, and that the corresponding absolute twist will then represent electric displacement, and the axis of twist is therefore, in free ether, in the same direction as the electric force. Hence it follows from the investigation on page 66 of the propagation of a plane wave that the electric and magnetic forces are both perpendicular to the direction of propagation, so that the vibrations are entirely transverse, that is to say, they have the character known to be possessed by light waves. The electromagnetic theory of light therefore accounts, in the case of plane polarised waves in free ether, for the hitherto unexplained absence of longitudinal vibrations from luminous waves.

Let u be the twist at any instant of a volume element of the ether situated at a point P, and let q be the ether displacement at P, that is to say, the displacement of P from its position of equilibrium, arising from the twist u. Then the relation between the vectors u and q may be expressed by the statement that $2u$ represents the *curl* of the vector q, or

$$u = \tfrac{1}{2} \text{ curl } q.$$

This equation may be taken as defining the curl of the vector q. As this concept will be of use later, it will be advisable to consider it a little further. In the illustration on page 67 I have considered the simple case of a distribution of twist in an incompressible medium such that the axis of twist is everywhere at right angles to the plane of the diagram, and the displacement,

of which twice the twist is the curl, is consequently everywhere in the plane of the diagram, and at right angles to the direction of propagation of the wave. That is to say, the direction of u, and consequently of q, remains constant, although the magnitudes of u and q are functions of the time, or of the co-ordinates, according to whether we are considering a sequence of changes in time, or the changes from point to point in space at a given instant.

The same equation may, however, be taken as the definition of the curl of a vector for any distribution of twist in an incompressible medium, or, expressing the definition in words:—

When an incompressible medium is distorted from its initial state by any distribution of twist, the twist at every point is said to be half the curl of the displacement at that point.[1]

It follows from the equation

$$u = \tfrac{1}{2} \text{ curl } q$$

that $\dfrac{du}{dt}$, or $\dot{u}$, the time rate of u, that is to say, the angular velocity of the twist of a volume element of ether, will be connected with $\dot{q}$, the time rate of ether displacement, that is to say, the velocity of ether flow corresponding to the twist, by the relation

$$\dot{u} = \tfrac{1}{2} \text{ curl } \dot{q}\,.$$

Now, suppose R to be the moment of the couple due to the elastic resistance of the ether to twist per unit volume, and we shall have

$$R = -ku,$$

where k is a constant, in the case of free ether.

In a non-isotropic material medium k will vary with the direction, being what we have before designated as a linear and vector function of a vector.

The potential energy W, and the kinetic energy T, in free ether, or in an isotropic material medium, will therefore be given by the expressions

$$W = \tfrac{1}{2}ku^2 = \tfrac{1}{2}k(u_1^2 + u_2^2 + u_3^2)$$

where u_1, u_2, u_3 are the component twists parallel to the co-ordinate axes; and

$$T = \tfrac{1}{2}\rho\dot{q}^2 = \tfrac{1}{2}\rho(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2)$$

[1] The mathematical reader will observe that this definition is identical with the usual definition of the curl of a vector q, viz., that the vector whose components are

$$\frac{\partial q_3}{\partial y} - \frac{\partial q_2}{\partial z}, \quad \frac{\partial q_1}{\partial z} - \frac{\partial q_3}{\partial x}, \quad \frac{\partial q_2}{\partial x} - \frac{\partial q_1}{\partial y},$$

is said to be the curl of the vector whose components are q_1, q_2, q_3.

where ρ is the density of the ether, and $\dot{q}_1$, $\dot{q}_2$, $\dot{q}_3$ are the components of ether flow parallel to the axes.

In *æolotropic* media, that is to say, media having different properties in different directions, such as crystals, for example, these expressions will assume the more general form

$$W = \tfrac{1}{2}(k_{11}u_1^2 + k_{22}u_2^2 + k_{33}u_3^2 + 2k_{12}u_1u_2 + 2k_{13}u_1u_3 + 2k_{23}u_2u_3),$$
$$T = \tfrac{1}{2}\rho(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 + 2\dot{q}_1\dot{q}_2 + 2\dot{q}_1\dot{q}_3 + 2\dot{q}_2\dot{q}_3).$$

If we identify electric displacement with ether twist by writing

$$u = A\mathbf{D},$$

where A is a constant, then the electrical energy must be identified with the potential energy W, and the magnetic energy with the kinetic energy T. The latter will represent the energy of ether flow.

This is the simplest method of expressing electric actions in terms of ether motions, and has been employed by Sommerfeld[1] and Larmor.

Another method, which has been employed by Boltzmann[2], would be to identify the electric displacement with ether flow, and the magnetic induction with ether twist. This is in accordance with Lord Kelvin's representation of magnetic induction as due to rotation of the volume elements of the ether, which has been very generally followed by other writers on the subject. This representation, however, leads to much greater difficulties when we endeavour to correlate it with the electron theory (see Appendix H). I have therefore preferred the former one.

It may be pointed out that neither of these representations is essential to the Faraday-Maxwell theory, either in its original form or in the development due to the introduction of the concept of the electron. All that is necessary is that, of the two vectors, magnetic induction and electric displacement, one should be proportional to the curl of the other, and we might content ourselves with the correlation of electrical phenomena as actions transmitted through the ether, without attempting to picture the method of transmission. We should then stop short of a true mechanical theory of electricity, that is to say, its expression in terms of the motion of some kind of substance, the ether, and we should have to abandon the attempt to correlate the electromagnetic scheme

[1] *Ann. Phys. Chem.*, xlvi., 1892.
[2] *Ibid.*, xlviii., 1893.

with phenomena such as gravitation and cohesion, which do not appear to find a place within it. We should also have to accept the definiteness of scale of material atoms as a fact not open to explanation, instead of attempting to account for it, as I have done later, by the correlation of the electromagnetic scheme with such actions; and finally, we should lose the great practical advantage offered by the simple and complete physical representation of electrical phenomena provided by the mechanical scheme in which electric displacement is ascribed to ether twist.

The result obtained above for plane waves may be extended to the general type of electromagnetic waves by the aid of Poynting's theorem, according to which the direction of the flow of energy is perpendicular both to the electric and magnetic forces. Now, the direction of the flow of energy at any point must necessarily be the direction of propagation at that point, that is to say, the path of a ray of light and the path followed by the energy must necessarily be identical; for the path of a ray may be defined as the locus of points at which the placing of a screen would arrest the ray, and the same definition must evidently apply to the path of the energy. The electromagnetic theory therefore affords a satisfactory general explanation of the absence of longtitudinal vibrations in luminous waves.

When the formulæ are applied to the explanation of the phenomena of double refraction in crystals it is found that, assuming the ether to be of the rotationally elastic M'Cullagh type, as we have shown good grounds for believing to be the case, the electric and magnetic forces must be respectively, in and perpendicular to, the plane of polarisation.

The boundary conditions are the same for every electromagnetic field, being simply the continuity of the tangential components of both the electric and magnetic force, and it will be shown in Chapter XII. that these conditions are always satisfied.

Maxwell demonstrated the existence of a radiation pressure in the direction of propagation of electromagnetic waves, the consequence of which is that luminous and other radiation exerts a mechanical pressure upon any material substances on which it impinges, but, as was pointed out earlier in the present chapter, his numerical results require qualification, owing to his neglecting the polarisation current.

As long ago as 1746 the great German mathematician Euler expressed the opinion that light waves must exert such a pressure, but the suggestion was not favourably received by his contemporaries, and the fact of its ever having been made was lost sight of for more than a century. The existence of the radiation pressure was demonstrated experimentally by the Russian physicist

Lebedeff in 1900, and his results were confirmed shortly afterwards in the United States by Nichols and Hull.

Maxwell points out that when an electromagnetic wave enters a conductor the disturbance will consist not only of electric displacements, but of currents of conduction, in which electric energy is transformed into heat, so that the undulation will be absorbed. Good conductors of electricity should therefore be opaque to light, while insulators should be transparent; and *vice versa*, opaque bodies should be good conductors, and transparent bodies should be insulators. This relation is found to hold good except in the case of electrolytic and gaseous conduction. Maxwell observes that:—

"Electrolytes allow an electric current to pass, and yet many of them are transparent. We may suppose, however, that in the case of the rapidly alternating forces which come into play during the propagation of light, the electromotive force acts for so short a time in one direction that it is unable to effect a complete separation between the combined molecules. When, during the other half of the vibration, the electromotive force acts in the opposite direction, it simply reverses what it did during the first half. There is thus no true conduction through the electrolyte, no loss of electric energy, and consequently no absorption of light."

If we interpret true conduction as conduction of the character occurring in metals, and omit the word combined, Maxwell's explanation will continue to hold good as far as it goes, in the light of our present knowledge. He notes the fact that the transparency of thin metallic films is very much greater than is consistent with the theory as developed in his equations, and this discrepancy is due to his neglecting to take into account the polarisation current, which becomes sensible in the phenomena of radiation. A very clear analysis is given of the observed facts of the magnetic rotation of the plane of polarisation discovered by Faraday, with a view, as Maxwell states, of preparing the way for an explanation, and as the electron theory provides this in a very complete and satisfactory form, it will be better to defer the consideration of this phenomenon until we are in a position to consider it from this point of view. I will therefore conclude the present chapter with the following important and suggestive paragraph from § 831 of Maxwell's treatise:—

"I think we have good evidence for the opinion that some phenomenon of rotation is going on in the magnetic field, that this rotation is performed by a great number of very small portions of matter, each rotating on its own axis, the axis being parallel to the direction of the magnetic force, and that the rotations of these different vortices are made to depend on one another by means of some kind of mechanism connecting them."

CHAPTER IX.

THE ELECTRON THEORY.

In considering Larmor's further development of the Faraday-Maxwell theory we shall adopt the simplification suggested in the preceding chapter, and regard the electromagnetic field as being due entirely to electric currents. Now, on the basis of the electron theory, which will be presently considered, it is found that when a magnetic distribution is specified in terms of a continuous distribution of electric currents, the electric and magnetic systems will be equivalent as regards magnetic induction, but not as regards magnetic force. In treating of a current system free from magnetism, the only magnetic quantity that occurs is the magnetic induction due to the currents, that portion of the induction which is represented by the expression $4\pi \mathbf{I}$ in the ordinary statical treatment of magnetism, and which forms the contribution of the part of the current arising from contiguous molecules, or elements of volume, being always negligible in comparison with the induction as a whole. Whenever, therefore, magnetic force and magnetic induction are identified in the preceding chapter, we shall consider the magnetic induction as the quantity dealt with. Magnetic force is, in fact, not one of the fundamental quantities of electrodynamic theory developed on the single basis of moving electrons, but is a concept introduced in passing from molecular dynamics to the mechanical theory in which individual irregularities are smoothed out by the consideration of the aggregates of molecules forming elements of volume and treated statistically. The magnetic force on this view will prove to be simply the average ethereal disturbance which is found to represent the magnetic induction, diminished, where necessary, by the part arising from purely local causes.

Larmor starts with the consideration of an ether of the M'Cullagh type at rest, and free from material molecules, and this part of the development is effected on the lines of M'Cullagh's analysis.

The kinetic energy per unit volume of the ether is first expressed

as half the product of a coefficient representing the mass or inertia of the ether, into the square of the time rate, or velocity, of a vector representing the displacement of the ether at any point arising from the strain due to an electromagnetic field. If this is compared with the expression for the electrokinetic energy obtained in the last chapter, bearing in mind that **H** is now to be replaced by **B**, it will be seen that the time rate of this vector may be taken to represent the magnetic induction. Now, the time rate of the ether displacement is simply the velocity of ether flow, so that on this view the magnetic induction at any point is represented by the velocity of ether flow at that point. The circulation round the tubes of magnetic induction therefore means simply an actual irrotational or twistless flow of the ether round the circuits formed by the tubes.

The potential energy per unit volume of the ether is then expressed as half the product of a coefficient representing the elasticity of the ether, into the square of a vector representing the absolute twist of the ether at any point, corresponding to the displacement which gives rise to the kinetic energy. The direction of this vector will be that of the axis about which the twist takes place, and its magnitude will be the measure of the angle of twist. If this expression is compared with that obtained for the electrostatic energy in the preceding chapter, it will be seen that this vector represents the ethereal portion of the electric displacement. The necessity for the coefficient depending on the elasticity of the ether has already been indicated by the presence of the factor $\frac{1}{c^2}$ in the formula for the electric displacement expressed in electromagnetic measure.

Tke kinetic energy T and the potential energy W of the whole field are then obtained by summations, extending throughout the whole of space, of the expressions for the energies per unit volume, and the conditions of equilibrium and equations of motion are then derived by the application of the principle of least action to the expression $T - W$. These equations are identical with M'Cullagh's optical equations, as well as with those obtained by Maxwell as expressing the electrostatic and electrodynamic actions in free ether.

Instead of proceeding at once to the formal definition of the electron and dealing with the results of its introduction, it will be more instructive to consider briefly the manner in which the concept was arrived at, and the first point to deal with will be the electrostatic attraction between material bodies.

When two electrically charged bodies are moved relatively to each other, the electrostatic energy due to ether twist will be

altered, but since this twist, constituting the electric displacement, fulfils the stream, or as Lord Kelvin calls it, the circuital condition, the electric charges will be unchanged in amount. Now, the ether has been shown to be necessarily a perfect fluid, that is to say, one which is entirely free from internal friction or viscosity, so that the loss or gain of energy can only be due to transference to or from some other system which is linked with the electric system, that is to say, connected with it by means of some kind of intervening mechanism. As a matter of fact, it is found to reappear in the form of mechanical energy of the charged conductors, and it is this mechanical energy which determines the system of pondero-motive, or mechanical, forces acting between them. In the absence of radiation, that is to say, of the excitation of electromagnetic waves, the whole of the overflow of electrokinetic energy produced by displacing a conductor in an electrically excited dielectric must display itself in ordinary mechanical forces acting on the conductor. The magnitudes of these forces have been experimentally measured in various media, and the results confirm this method of accounting for their origin, and appear to be in accordance with those arrived at theoretically, as indicated in the preceding chapter. When a dielectric is moved through an electric field, a similar transformation of energy takes place at the boundary between the two dielectrics.

The mechanical traction along the normal to the boundary, or interface, when it is distributed over the surfaces of a pair of conductors separated by an excited dielectric, as in the case of the two coatings of a Leyden-jar, may be mechanically balanced by a stress in the substance of the dielectric. This is the only kind of mechanical stress in a dielectric of which we can obtain experimental evidence. Korteweg, H. von Helmholtz, Kirchhoff, and others, have calculated its value in special cases, and verified the existence of such strains by the experimental determination of the changes of volume of the dielectrics under electrification. The quantitative results of these *electrostriction* phenomena, as they are called, cannot be relied on, as Larmor showed in 1897. The reason for this may be seen in a general way in the fact that any inequalities of twist in contiguous portions of a solid dielectric will necessarily give rise to secondary strains which will mask the effects directly due to the electrodynamic forces. These experiments, however, and others in which strains of this character have been shown capable of transforming an isotropic transparent dielectric into a doubly refracting one, afford independent evidence of the existence of such strains.

In the case of the intervening dielectric consisting of a fluid, as air, for example, the mechanical traction on the conductors may

be balanced, and therefore measured, by means of weights, as is done in some of Lord Kelvin's electrostatic measuring instruments.

In addition to this normal traction, the theory shows that on each element of interface in the dielectric ether there will be a purely tangential traction at right angles to the tangential component of the electric force, and equal to it in amount. The existence of these stresses cannot be directly demonstrated experimentally, since we have no means of making direct observations on the ether.

In order to deal with the electrodynamic problem, it is necessary to devise some simple form of circuit in which a continuous electric current can be maintained without the introduction of voltaic cells or other apparatus which could not easily be included in the expression for the energy. This may be effected by means of a condenser with its pair of coatings connected by a thin wire, as

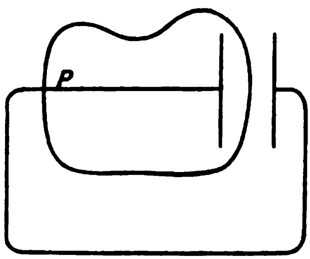


FIG. 11.

shown in fig. 11, the two plates of the condenser being supposed to be steadily moved towards each other while the coatings are thus connected. To fix the ideas, suppose a thin, positively charged, sheet of solid dielectric to be placed between the two plates, and in contact with the right-hand plate, the remainder of the interspace being occupied by air. There will then be an accumulation of negative electricity on the right-hand plate, and of positive electricity on the left-hand one. The approach of the plates at a uniform speed will then produce a steady current in the conducting wire, which will flow round the circuit from the left-hand plate to the right-hand one. The only variations from linearity in such a circuit will be in the condenser itself, and it may be made negligible by making the dielectric plate sufficiently thin. Supposing the resistance of the wire to be negligible, a steady current in a conductor devoid of resistance can be realised in this manner, without the aid of any complicated source of electromotive energy.

We now have to see how this steady current can be accounted

for on a purely dynamical basis. We must note, in the first place, that the motion is very slow in comparison with the speed of electric radiation. From this it will follow that the interior of the dielectric will be, at any instant, sensibly in a condition of equilibrium, for the same kind of reason that the slow backward and forward motion of a body does not give rise to appreciable sound waves. The result of this, as is shown in the development of the conditions of equilibrium by the application of the principle of least action, is that the twist vector which we have identified with the electric displacement is derived from a potential V, and at the surface of any of the conductors, supposed to be practically perfect, that is to say, of insensible resistance, is in the direction of the normal, provided the medium is isotropic. In other words, it is, at each instant, the electric displacement determined by the charges existing in equilibrium on the faces of the condenser and on the connecting wire. Owing to the condensing action, this electric displacement is small relatively to the charges, except between the condenser coatings and close to the wire. Let any closed surface be described passing between the plates of the condenser and intersecting the wire at the point P. The medium is now discontinuous at the surface of the wire, owing to the fact that electric displacement cannot be sustained within the wire. We will now suppose the elastic continuity of the medium inside the conductors to be restored, while such a flow of electric displacement is imparted along the wire as will leave the state of the field unaltered, so that there will be no disturbance within the conductors. The electric displacement, being given as a twist, must, from its rotational nature, have the circuital or stream character, and therefore its total flux through the closed surface will be zero. The flow of electric displacement must therefore be such that its flux through the closed surface remains zero when the plates of the condenser are slightly approached towards each other. This movement will, however, cause an increased concentration of the already large proportion of the flux passing across that portion of the surface lying between the plates, and, since the flux through the remainder of the closed surface, with the exception of the part near the wire, is insignificant, this concentration will have to be balanced by a very considerable increase of the imparted flux close to the surface of the wire, or along its surface if the wire be regarded as a perfect conductor. When this change in the capacity of the condenser ceases, the electric displacement will at once resume the condition of equilibrium in which it is everywhere directed along the normal to the surface of the wire. This intense flux, close to, or along, the surface of the wire, continuing as long as the capacity of the condenser is changing, is simply the

current in the wire, since the flux is the electric displacement in the field; and the whole of its circumstances have here been deduced from purely dynamical considerations.

If we suppose the condenser to be suddenly removed, and the wire made continuous, while the current is flowing round the circuit, there will remain an ordinary circuital current which would continue to flow without alteration if the conductor were completely without resistance. Such a permanently circulating current might be considered as a vortex ring in the perfect fluid constituting the ether, and would have the important distinction from an electric current flowing in an ordinary circuit, in that it would remain constant throughout all time, being entirely unaffected by electromagnetic induction. In the condenser circuit, on the other hand, the current depends on the rate of relative movement of the condenser plates, that is to say, it is affected by changes in the twist strain energy of the portion of the medium which occupies the gap in the conducting circuit. Any relative motion of the condenser plates gives rise to a flow of electric displacement across a closed surface which passes between them, and must therefore be taken as giving rise to an equal and opposite flow where this surface intersects the connecting circuit. This flow, or current, representing the action of the channel of discontinuity on the elastic transmission in the medium, implies a circulation of the fluid medium completely round the conducting circuit, and it is this flow which provides the kinetic energy of the electric current. A current in a conductor has no sensible elastic potential energy for the simple reason that, in the case of movements of ordinary velocity, the medium is always sensibly in a state of equilibrium, any beginning of an electromotive disturbance of the steady motion being equalised before it has time to grow. A complete circuital current, as represented by the perfectly conducting circuit after the removal of the condenser, will therefore maintain its strength unaltered, the surrounding ether moving in such a manner that the electro-dynamic induction in the circuit is always null. When, however, the current circuits are completed across the dielectric, or through an electrolytic medium, as when a voltaic cell forms part of the circuit, the constraint to nullity of induction is removed, and there will no longer be constant circulation.

The complete circuital current may be considered as representing a permanent magnetic element according to Ampère's theory. Such an element might therefore, on this view, be regarded as a circuital cavity or channel in the elastic ether, along the surface of which there is a distribution of twist, or *vorticity*, as it is called. That is to say, a permanent magnetic element would be represented by a vortex ring with either a vacuum or a portion of the

fluid ether, devoid of elasticity, for its core. This would make the magnetic induction, as we have taken it to be, consist of a flow of ether through the cores of a mass of such elements so oriented as to give rise to a continuous circulation in one direction round a magnetic circuit, whether the circuit be completed in magnetic matter or contain gaps filled either with free ether or with material dielectrics.

In order to account for magnetism in this manner, such a vortex ring must be supposed to form a part of the constitution of every molecule of magnetic matter, and as all kinds of matter are found to show traces of magnetic properties, the conception must be extended to all material molecules.

Such vortex rings, being permanent structures in the frictionless ether, could not be supposed to be broken up at the temperature at which iron, for example, loses its magnetic properties, to reappear on lowering the temperature, so that a permanent electric current of this kind would have to be involved in the constitution not only of every molecule, but of every material atom. It might be supposed that the atoms of a strongly magnetic substance, such as iron, grouped themselves into aggregates with their atomic currents so directed as not absolutely to oppose each other's action, and that at higher temperatures these groups were broken up and replaced by groups of a different type, for each of which the actions at a distance of the atomic currents were mutually destructive. Combinations of the atoms into molecules of the latter type would account for the existence of what are commonly called non-magnetic substances, and this theory would account in a perfectly satisfactory manner for the extremely small magnetic moments of the molecules of such substances.

These considerations lead to a representation of material atoms consisting, as was suggested by Lord Kelvin, of vortex rings in the ether, and this suggestion formed a most important step towards a consistent representation of physical phenomena. An objection raised by Maxwell against the theory was that it did not account for the mass, or inertia, of matter; and we shall see presently that it has other deficiencies which led Larmor to the introduction of the concept of the free electron, leading to an amplification of the vortex ring theory of material atoms, in which these deficiencies are obviated. The easiest path to a clear understanding of the amplified theory is by the prior consideration of the consequences of the simpler theory of Lord Kelvin.

Some very beautiful and interesting experiments may be made with vortex rings in air, the rings being rendered visible by means of smoke, or other visible clouds of solid particles. A simple method is to make a wood or cardboard box of any convenient size, having

a circular aperture cut in the middle of one of its sides, and having the opposite side, or a part of it, hinged, so that it can be opened and suddenly closed. The box may be filled with smoke, or preferably with a visible cloud of particles obtained by soaking some absorbent substances, such as pieces of blotting paper, separately in solution of ammonia and in hydrochloric acid, and placing them side by side on the bottom of the box, which will soon become filled with a white cloud consisting of solid particles of ammonium chloride.

A ring projected, by suddenly closing the hinged side of the box, in such a manner that its coreless centre impinges on the flame of a candle, will blow the latter out, even at a distance of 30 or 40 feet or more. The reason of this is to be found in the method by which the ring is formed. The sudden compression of

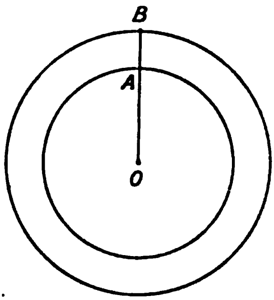


Fig. 12.

the air in the box forces some of it out through the circular aperture, and the circumferential portions, through friction with the circular boundary of the aperture, are retarded relatively to the more central portions. This gives rise to a forward moving ring of air shaped like an anchor ring, so that any section AB of the ring, made by a plane passing through the centre O (fig. 12) at right angles to the plane of the ring, and therefore to the plane of the aperture, is circular. The point A will be moving forward more rapidly than B, because less retarded, so that every circular section AB will travel like a wheel turning about an axis in the plane of the paper, and perpendicular to AB. The result of this will be that the air inside the wheel will be dragged forward by friction with the rolling portions, and the current of air produced in this way will blow out the candle as the centre O passes through the flame.

A ring of this character in a frictionless fluid, which, it must be remembered, could neither be created nor destroyed by any

method conceivable to us, could not be cut through by the sharpest knife, but would alter its shape as the knife approached it. Two rings approaching each other would rebound elastically as if made of indiarubber. If the lines of motion were oppositely directed in the same straight line, the rings would increase in diameter and diminish in cross section as they came near together, and before rebounding, and the same thing would happen in the case of a ring approaching a plane surface perpendicular to its direction of motion. If a ring were made to follow a more slowly moving one, with the centres of both travelling in the same direction along the same line, the second ring, on approaching the first, would become smaller and pass through the first. It would then begin to increase in size and diminish in speed until the first one in its turn had passed through the second, and so on. These results can, of course, only be completely demonstrated mathematically, but they may be partially imitated in the case of the transient rings obtained as described above, and the experiments are both interesting and instructive, as these considerations show that aggregates of vortex rings would possess many of the properties of ordinary material substances.

The mathematical theory of rotational or vortical fluid motion, as developed by von Helmholtz, Lord Kelvin, Sir J. J. Thomson, and others, shows that such vortices in a perfect fluid would unite to form molecules and molecular groups. The vortices in each group would, however, tend to form aggregates similar to those formed by elementary magnets, in which the separate elements would reinforce, instead of neutralising, one another's effects, when placed in a magnetic field. On this theory, therefore, substances ought to be about equally magnetic at all temperatures, unless there be some other bond between the atoms of the molecule, in addition to the one due to the mutual actions of the vortices exerted through the fluid in which they are formed; and this additional bond would have to be of about the same order of magnitude.

An additional bond is afforded by the electric charges of the atoms required by the theory of electrolysis, but even then, unless each molecule consists of more than two atoms, arranged in some symmetrical order, about half of the molecules must be so constituted as to reinforce each other's effects.

This theory, as it stands, would not admit of Weber's explanation of diamagnetic action as being due to induced currents in the molecules owing to the action of an external magnetic field, for, according to this view, the molecular currents are completely circuital, and therefore cannot be altered by any external action.

We shall also see presently that the vortex atom theory does

not, in itself, provide any means for the communication of an electric charge to a conductor, nor does it afford any indication of how electric charges can be associated with the vortex atoms.

The complete circuitality of the molecular currents, which is a necessary consequence of this representation of atoms as simple vortices, leads to the difficulty that the only explanation which it can possibly give of the actions between permanent magnets is one which is inconsistent with experimental results.

Lord Kelvin, in 1870, solved the problem of the mutual actions of a series of vortices which, instead of having vacuous cores, through which the free circulation of fluid corresponding to magnetic induction takes place, have open rigid tubes as cores, with free circulation through these tubes. The force system, or forcive, in this case is found to be equal in amount, but *opposite* in direction to the forcive between the corresponding steady electric currents, while the forcive between permanent magnets is found experimentally to be in the *same* direction as that between the corresponding currents. It is unnecessary to inquire whether any physical meaning could be ascribed to such rigid tubes, as it can be proved in a perfectly general way, and without making any special hypotheses, that any forcive acting between such unalterable circuital currents must be in a direction opposed to that acting between ordinary electric currents. In the latter case the currents are susceptible to alteration by magnetic induction, so that interconnections must exist between the current system and the external field, in order that the varying twist in the external ether may be communicated to the system of circuits. In the former case, on the other hand, the circuits form channels of purely cyclic motion entirely unconnected with the external field. This *cyclosis*, as it is called, when existing in any field of energy, entirely changes the type of the equations derived from the principle of least action, and leads to the result above stated.

The modification which must be introduced into the vortex atom theory, in order to obviate this discrepancy, must therefore be of such a character as to give the elasticity of the medium some kind of grip upon the atomic circuits, and so destroy their purely cyclic character, without, however, destroying their permanence. This may be effected by supposing the vortex rings to be provided with cores composed of discrete electric nuclei, or centres of radial twist, in the medium, and to suppose the currents to be currents of convection due to the circulation of these nuclei around the circuit of the core. Such a vortex as this would be capable of moving about in the medium without experiencing any forces on the circulating nuclei which would tend to break it up. The hydrodynamical stability due to its vorticity, or spin, would be

sufficient for its permanent maintenance. The circulation forming the atomic current will, however, now be subject to variation owing to its elastic connection with the medium, so that it will no longer have an independent, purely cyclic motion. A magnetic atom of this type would therefore behave like an ordinary electric current in a non-dissipative, that is to say, a perfectly conducting, circuit, so that the atomic currents would be capable of variation due to external magnetic induction, and Weber's explanation of diamagnetism would hold good. The formation of the electric doublets, out of which such a magnetic molecule may be built up, is illustrated by Larmor in the following manner:—Let a narrow tubular channel of molecular dimensions be made in the rotationally elastic ether, as illustrated in Lord Kelvin's model (p. 71), and then suppose the walls of this channel to be grasped, and twisted round the axis of the tube: this twist will be distributed through the medium, and, as a result, there will be lines of twist displacement, all starting from one extremity of the tube and terminating at the other extremity. As long as the walls of the tube are held in this position by external constraint, one extremity will constitute a positive electron in the medium, while the other will be the complementary negative one. If the walls of the channel are released, both will disappear together. Now suppose that, before this release takes place, the tube is filled up with ether constituted as we have imagined, with the exception of small vacuous nuclei at the extremities, which nuclei will then assume the spherical form. If the external constraint is then removed, the effort of release in the surrounding medium will twist the spheres within the tube to a slight extent until a condition of equilibrium is attained in which the rotational elasticity of the medium inside balances that of the medium outside. Each doublet will then consist simply of a self-locked intrinsic strain in the medium, analagous to that existing in a material wire which has been welded into a ring after a twist has been put upon it. A positive electron will then differ from a negative one essentially as the reflection of an object in a plane mirror differs from the object itself, but it is not necessary that the tube should be cylindrical in section and therefore the vacuous extremities equal in size. Evidence has been given in Chapter VII. of the existence of negative electrons of extremely minute size and having the same properties whatever the nature of the material from which they were derived. The positive carriers, on the other hand, were found to be comparatively large and heavy, and appear to consist invariably of portions having the properties of the material from which they were derived, and but little evidence has yet been obtained of the independent existence of positive

electrons. We have already obtained evidence of the atom being an aggregation of simpler systems, and we have now to consider whether it be possible to form an aggregate of electron doublets which will agree in its behaviour with the properties known to be possessed by material atoms. This is found to be more easily done on the assumption that the positive electron occupies a space which is large in comparison with that occupied by the negative electron. In order that an electron should move through such a medium without giving rise to twist, the inertia of the latter, and therefore the moments of momentum of the moving spheres in the Kelvin model, must be supposed large. It is also necessary to assume a definite law of twist along the tube joining a positive electron to a negative one; for example, if the tube were cylindrical, the twist would have to be constant all along its length, whether that were straight or curved. Under these conditions an electron would, on the assumption that the ether displacement of the nucleus is negligible, a point which will be presently considered, move through the medium without experiencing any resistance,[1] and therefore a tube connecting a pair of complementary electrons would experience no resistance in moving through the medium in the direction of its length. If, however, such a tube were made to move sideways through the medium, it would give rise to twist, and would therefore tend to drag some of the medium with it. Its motion would therefore be opposed. If we suppose, with Sir J. J. Thomson, that there is a resulting bodily drag of the ether, this opposition may be represented as arising from the inertia, or mass, of the ether thus carried along with the electron, and this affords a very simple explanation of the mass, or inertia, of material bodies.

Another analogy suggested by Larmor is that of a flexible shaft, such as is sometimes employed to transmit rotation from one line of shafting to another line inclined at an angle to the first. Suppose that the ideal channel in the ether, instead of being filled up by a filament of ether, were filled up by such a shaft, or wire, of indefinitely great torsional rigidity, and in continuous connection with the surrounding ether. Whenever any cross section of this wire is turned round its axis, the resulting turn, or twist, will be transmitted all along the wire, and thence to the ether alongside it, forming a pair of complementary electrons at its ends. On releasing this section the twist will undo itself, and the terminal electrons will disappear. This arrangement forms an elastic system devoid of the intrinsic stress previously implanted by filling the canal with ether; for it becomes free from stress on

[1] The question of the mobility through the ether of electrons so constituted is discussed mathematically in Appendix H.

releasing the wire. We are therefore able to distinguish the proximate cause of the attraction of one electron for the other, as arising from the tangential tractions exerted by the surrounding ether on the surface of the wire. These tractions, by the principle of the conservation of energy, must constitute a system of forces equivalent to an attraction between the ends. The ether with its contained electrons might in this way be imagined to be mathematically dissected into a system devoid of intrinsic strain, by connecting each positive electron with a complementary negative one by means of such an elastic material wire in continuous connection with the ether, assuming that the appropriate twist has been imparted to each wire in order to maintain the intensities of the electrons. When the wires have been removed, and the electrons have been permanently constituted by filling up their places with ether, the possibility of thus specifying a proximate cause of the mechanical attraction between the electrons may also be considered, in a sense, to have disappeared, for only the ends of the tubes will then be determinate. We may still consider the tractions exerted on the surface of such a tube of strain as statically transmitted to the electrons at its ends, just as if it included the wire of our illustration, but the manner in which these electrons are to be cross connected will no longer be a definite one.

Since, as mentioned in Chapter I., Sir J. J. Thomson has shown that a moving body carrying an electric charge possesses a definite inertia owing to the charge, and has determined its amount in the case of a charge uniformly distributed over the surface of a sphere of definite size moving at a definite speed, it follows that the mass, or inertia, of matter would be accounted for by any theory in which an electric charge forms an essential constituent of the material atom.

The tube joining two complementary electrons is clearly a tube of electric induction, or what we have called simply a Faraday tube. As the electron forms a natural unit of electric quantity, we may define a unit Faraday tube as the tube of electric induction joining two complementary electrons, and we see that the Faraday tubes must not be considered as forming an ultimately continuous distribution, but as having an actual discrete existence, giving a fibrous structure to the ether. We shall see later that, even when most closely packed, the space occupied by the tubes will be small in comparison with the space free from them. The two electrons connected by a Faraday tube may be separated to any distance; but in order to do this, work will have to be done in extending the state of twist surrounding the tube through the ether, and if the constraint is removed, this twist will tend to come out again, causing the electrons to approach until the whole

tube has shrunk to its original molecular dimensions. This will happen if the dielectric separating the electrons is replaced by a conductor.

Since the twist in the ether surrounding a Faraday tube is in a plane perpendicular to the axis of the tube, and is reversed in direction by reversing the direction of the tube, it follows that the approach of two similarly directed tubes will increase the surrounding ether strain, while the approach of two oppositely directed tubes will tend to make them counteract each other's effects, and so diminish the strain. If free to move, the tubes will, of course, move so as to ease the surrounding strain, so that similarly directed tubes will repel one another. Two oppositely directed tubes, on the other hand, will attract one another, so that, if free to move, they will approach as nearly as possible.

In this way we arrive at a result corresponding very closely with Faraday's original conception of actual physical lines of electric force resembling bundles of stretched, mutually repellent, springs extending throughout an electric field. The lines of electric force, which we have generalised into lines of electric induction, may, in a sense, be considered to have an actual physical existence, in that they map out lines of actual physical twist, but we must guard ourselves against supposing that this implies that the same identical pair of electrons always remain bound together, as they would be compelled to if connected by extensible material springs. Every electron must have its complementary partner, but there is nothing to prevent a change of partners taking place under suitable circumstances, although every electron must be the terminal of a unit Faraday tube. The lines of strain between the electrons of a system will therefore in all cases give the ether the fibrous structure referred to above. The importance of this was pointed out by Sir J. J. Thomson in his very interesting Silliman lectures on Electricity and Matter, after first calling attention to one of those extraordinary examples of intuition or prescience, whichever we like to call it, which we find so frequently in Faraday's writings. In his paper entitled "Thoughts on Ray Vibrations," Faraday says, "The view which I am so bold to put forward considers therefore radiation as a high species of vibration in the lines of force which are known to connect particles and also masses together." It is precisely by the vibratory motion of these tubes that, according to our present theory, all electromagnetic waves are excited.

Sir J. J. Thomson points out that this point of view leads to the conclusion that electromagnetic waves, including such special cases as waves of light and Röntgen rays, must have a structure. The front of a light wave will therefore, instead of

being uniformly illuminated, be represented by a series of bright specks on a dark background, the bright specks occurring wherever the wave front is intersected by the Faraday tubes. Thomson suggests that such a discrete structure of the Röntgen ray pulses would afford an explanation of the comparatively small ionisation effected by such rays in their passage through a gas. He observes that such rays may pass very long distances through the gas without, even in the case of the strongest rays, ionising more than a billionth of the whole number of molecules contained in the gas. If the wave front were continuous, so as to expose the whole of the molecules to the ionising action of the rays, this small effect would be incomprehensible, except on the supposition that the only molecules capable of being split up were a small proportion characterised by some special condition, which might be, for example, the possession of an energy of motion very greatly exceeding the average kinetic energy of the molecules of the gas. If this, however, were the explanation, the ionisation ought to increase rapidly as the temperature of the gas rises. M'Clung has shown that no appreciable increase in the ionisation is produced by raising the temperature of a gas from 15° C. up to 200° C. On the hypothesis of the discrete structure suggested, the phenomenon would be easily explicable, as only a small proportion of the molecules might be expected, under such circumstances, to come into collision with the rays.

The effect of moving Faraday tubes may be very simply illustrated by what occurs when two parallel plates, one charged positively and the other negatively, are connected by means of a metallic wire. Before the introduction of the wire, the Faraday tubes will be cylinders having their axis perpendicular to the plates, each tube being held in equilibrium by the repulsion of surrounding tubes. All round the edges of the plates, the tubes being no longer held in equilibrium in this manner, will bulge outwards, forming a sort of fringe round the boundary of the space included between the plates. When a wire is led from a point on one plate to a point on the other, it will necessarily cut through some of the outlying tubes. Now, when a tube enters the substance of a conductor, the constraint which maintains it in extension is removed, and the tension in it causes it to contract to molecular dimensions, so that the repulsion which is exerted on the neighbouring tubes disappears. The result will be that more and more tubes will be pushed so as to intersect the area partially enclosed by the plates and the wire, causing a movement of the whole set of tubes in this direction. The discharge of the pair of plates will therefore be accompanied by a motion of the tubes at right angles to themselves. Now we know that the

physical effects accompanying this motion are an electrical current flowing from one plate to the other through the wire, and the production of a magnetic induction which, if the wire is in the form of a plane curve, will act along a line perpendicular to that plane, and the direction of the force along this line will depend on the direction of the current.

When a Faraday tube is in motion through the ether at right angles to itself the change of stress, depending on the speed, will alter the twist round the axis of the tube, constituting an electric displacement along the axis, and an ether flow, constituting magnetic induction at right angles to the axis and to the direction of motion, as shown in fig. 13.[1] The momentum must be in the

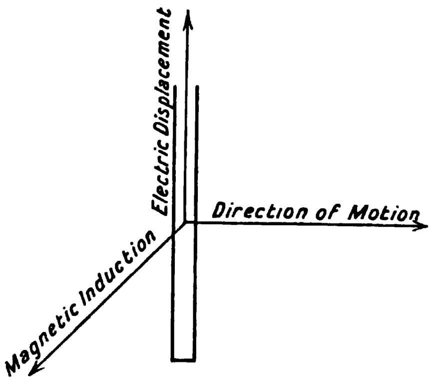


FIG. 13.

direction of the flow of energy, that is, by Poynting's theorem, at right angles to the magnetic induction and the electric displacement, viz. in the direction of motion. The electric force, being at right angles to the direction of motion and to the magnetic induction, will keep the tube at right angles to the direction of motion in an isotropic medium such as the ether, and in the absence of electrification, and therefore of electric force other than that due to the motion of the tube itself.

In accounting for the effects of the motion of small charged bodies such as electrons, or even atoms, the simplest supposition

[1] Professor Minchin has pointed out in a letter to *Nature* that, since the induction, the current (corresponding to electric displacement) and the direction of motion are represented in order by the directions of the Hamiltonian vectors i, j, k, as usually drawn, a convenient *memoria technica* for remembering their relations is given by the words "induction makes kurrent jump."

that can be made is that the carriers are spheres. It is easily shown that when a charged sphere is moving with a velocity which is small compared with that of light, the Faraday tubes will be uniformly distributed, and will all be directed along the radii of the sphere exactly as if the sphere were at rest; and it can also be shown that the energy of motion of the sphere will be increased by the presence of the charge on it in a manner which can be represented by a certain addition to its mass, so that part of the mass of the charged sphere must be considered as being due to its charge. Here we have an example of the motion of Faraday tubes at right angles to themselves. It was mentioned earlier in the chapter that when a tube moves in this way it drags some of the ether with it, and the added mass of the sphere due to its charge may be considered as the mass of ether dragged along by the tubes belonging to the sphere. A sphere moving through an ordinary liquid drags some of the liquid with it, and Green showed in 1833, in a paper on the Vibration of Pendulums in Fluid Media, that the sphere behaves exactly as if its mass were increased by that of a volume of the liquid equal to half the volume of the sphere. Later investigations have shown that when a cylinder moves at right angles to itself its mass is increased by the mass of an equal volume of the liquid, but this apparent increase of mass is very much smaller when the cylinder moves in the direction of its length. The amount of ether carried along by a bundle of Faraday tubes depends, not only on the number of tubes in the bundle, but on the closeness with which they are crowded together; and it is found that the quantity of ether thus *bound* in a given volume of a field of electric force is proportional to the square of the number of unit Faraday tubes which pass through that volume. Let us now return for a moment to our material analogy, and consider how the long, narrow cylinder forming the analogue of a Faraday tube would behave when moving through a liquid. It can be shown from general hydrodynamical considerations that such a body, if free to twist in any direction, would not move end on, but would set itself broadside to the direction of motion, placing itself, in fact, in such a position as to carry with it the greatest possible amount of liquid. A familiar example of this principle is seen in the case of falling leaves, which flutter down with their planes more or less horizontal, instead of falling edge first. Another illustration is afforded by the manner in which a boat, when being towed, tends to set itself across the line of motion, a tendency which has to be counteracted by careful steering, in the case of vessels of any size. When there is no one steering, as, for example, in the case of a punt towed behind a yacht, the punt will be seen to sway

from side to side in its attempts to set itself broadside on to the direction of motion, with the result of considerably increasing the towage resistance.

The application of this principle to the case of the moving charged sphere shows that the Faraday tubes belonging to the sphere will tend to set themselves at right angles to the direction of motion. The result of this tendency would be to crowd all the tubes into the plane passing through the centre of the sphere at right angles to the direction of the motion; but this is counter acted by the mutual repulsion of the tubes. Oliver Heaviside showed in 1889 that if V is the velocity of light in the medium, and v is the velocity of the moving sphere, then the distance of every point in the tube from this plane will be reduced in the proportion of V to $\sqrt{(V^2 - v^2)}$.

This result shows that the change in the distribution of the Faraday tubes becomes appreciable only when the velocity of the charged body approaches that of light. As long as this is not the case, its mass will therefore remain sensibly constant, but when its velocity approaches that of light, the mass will rapidly increase; and when the velocity becomes equal to that of light, the mass will become practically infinite, which means that some catastrophe must occur, probably the disintegration of the moving body.

These results indicate the possibility that the property of matter which we call mass or inertia may be entirely electrical, and some very interesting investigations have been carried out by Kaufmann (*Comptes Rendus*, Oct. 1902) with the object of determining this question. He found that if the charged body behaved like a metal sphere, the electrical mass would be about one-fourth of the whole mass, the ratio varying with the assumptions made as to the form of the moving body and as to whether it is taken to be insulating or conducting. Sir J. J. Thomson has calculated the ratio of the masses of the rapidly moving particles emitted by radium to the masses of the same particles when at rest or moving slowly, on the assumption that the distribution of the tubes of force is the same as of those due to a charged point, so long as we confine our attention to the field outside a very small sphere having its centre at that point, and on the further assumption that the whole of the mass is electrical. These results agree very closely with those obtained experimentally by Kaufmann, and justify the conclusion that the motion of an electron through the ether may be represented, to a close approximation, by that of a small material sphere carrying an electric charge. It must, however, be borne in mind that such a representation cannot in any sense be regarded as a possible picture of what is actually taking place in the ether. An electron

is essentially a strain form in the ether, that is to say, it is an ether structure. In considering the properties of such a medium, which can be directly expressed in abstract terms only, material analogies, such as Lord Kelvin's model of the ether, and the Larmor electron expressed in terms of that model, are of great assistance in fixing our ideas. We must, however, never forget that the ether is not a structure built up out of material elements, but is in all probability the primeval basis of which matter itself is formed. Larmor's conception of the electron, although expressed in terms of the material model, greatly assists us in thinking of a strain form in the ether. If we picture it first simply in terms of the material model, and then allow the material elements to dissolve, as it were, and disappear as completely as possible from our minds, we shall succeed in forming a much more definite idea of a strain form in the ether than would have been possible without the aid of the material illustration. The representation of an electron as a small material sphere carrying an electric charge is of a totally different character. It does not help us in the least towards a mental representation of an electric charge as a strain form in the ether to consider the action of an electric charge carried by a moving sphere—a conception of a totally different character. But although this representation affords no immediate assistance in the direction of picturing the nature of an electron, it has the great advantage of being amenable to exact mathematical development, and Kaufmann's observations warrant the conclusion that the results of calculations arrived at in this manner will be approximately correct. The reason for this is not that a sphere carrying an electric charge is even a possible approximation to a physical representation of an electron. What Kaufmann's observations show is that the actions which would be called into play by the motion of such a charged sphere are subject to approximately the same mathematical relations as those arising from the motion of an existing electron; but we must never forget the crudity of the representation, nor be discouraged if it should sometimes lead us into difficulties.

These considerations appear to confirm the validity of the assumption that the inertia of the electron arising from the ether carried along by its nucleus when in motion is negligible in comparison with that which arises from the drag of ether due to the motion of the charge, that is to say, of the strain distribution. The latter arises from the ether twist, and therefore constitutes the electromagnetic mass. The portion due to the ether displaced by the nucleus corresponds to a twistless ether flow, and although ethereal in origin, is not, I think, properly described as electromagnetic. From this point of view, the dimensions of the electron,

regarded as a small spherical body, must be regarded as determined almost wholly by the strain distribution, and therefore it would follow that the inertia of an electron considered as a sphere moving through the perfectly fluid ether should be of the same order of magnitude as the electromagnetic inertia of a charged sphere of the same dimensions. This has been shown to be the case in a remarkable paper by Sir Oliver Lodge on the Density of the Ether (*Phil. Mag.*, April 1907).

In the course of his introductory remarks, Lodge observes that:—

"Just as the ratio of mass to volume is small in the case of a solar system, or a nebula, or a cobweb, I have been driven to think that the observed mechanical density of matter is probably an excessively small fraction of the total density of the substance, or ether, contained in the space which it thus partially occupies—the substance of which it may hypothetically be held to be composed. Thus, for instance, consider a mass of platinum, and assume that its atoms are composed of electrons, or of some structures not wholly dissimilar; the space which these bodies actually fill, as compared with the whole space which in a sense they 'occupy,' is comparable to 10^{-10}th of the whole, even inside each atom; and the fraction is still smaller if it refers to the visible mass. So that a kind of minimum estimate of ethereal density, on this basis, would be something like ten thousand million times that of platinum."

These considerations are in accordance with Sir J. J. Thomson's conclusion that the mass of an electron is to be attributed to the "bound ether" which it carries with it as it moves through the ethereal medium, and we have seen that the motion of a sphere through an incompressible fluid would be the same as that of a sphere moving through space without resistance, but with an increased mass, the addition to the original mass being the mass of the amount of fluid which would fill half the space occupied by the original sphere. Now consider the electron as a sphere of radius a, moving through the sensibly incompressible fluid ether of density ρ, which it simply displaces. Then, the ether being incompressible, the simplest hypothesis that can be made is to suppose the density inside and outside the sphere to be the same. The mass of the sphere will then be the product of its volume, $\frac{4}{3}\pi a^3$, by the density of the ether, that is to say, $\frac{4}{3}\pi\rho a^3$.

Increasing this by half its amount, we obtain the value $2\pi\rho a^3$ as the effective mass of the sphere moving without resistance. If we knew the value of ρ, we might consider this as a second approximation to the inertia of the spherical electron, the value $\frac{4}{3}\pi\rho a^3$, that is to say, the mass, in the ordinary mechanical interpretation, of its volume of ether being taken as the first approximation.

A more accurate method of estimating the inertia of the

electron will be to calculate it directly from the magnetic force which it establishes and carries with it in its motion.

Now Sir J. J. Thomson has shown (*Recent Researches*, p. 13) that, as a consequence of Maxwell's equations, the magnetic force produced at a point P by a Faraday tube moving with a velocity v at P is $4\pi v \sin\theta$, where θ is the angle between the axis of the tube and the direction in which it is moving. Let e be the charge of the electron, moving with the velocity v, supposed to be small in comparison with that of radiation; then the Faraday tubes emanating from the electron will be uniformly distributed over the surface of the sphere, and will all be radial in direction. Let the straight line OP, joining the centre of the sphere to the point P, have the length r, and make an angle θ with the direction of motion. The density of the Faraday tubes at P, that is to say, the number of tubes per unit of area of a sphere with centre O and radius r, will then be $\frac{e}{4\pi r^2}$, and therefore the magnitude of the magnetic force H at P will be given by the equation

$$\mathrm{H} = \frac{ev \sin\theta}{r^2},$$

and its direction will be perpendicular to OP, and to the direction of motion.

Lodge points out that—

"we may here observe, without any hypothesis, how enormous is the magnetic intensity immediately surrounding the equator of an electron, moving along an axis at, say $\frac{1}{30}$th the speed of light. It is no less than 10^{15} C.G.S. units. At a third of the speed of light, even in the completer theory of which the above is the dominant term at ordinary speeds, it is still not appreciably more than 10^{16}; but it begins to run up suddenly towards infinity when the velocity of light is closely approached, the reason probably being that it is then no longer a small fraction of the intrinsic rotational energy of the ether itself."

Now, if the velocity of ether flow in the tube of induction at the position where the magnetic force or induction is H, be represented by w, then the energy per unit volume of the ether at this point will be represented by either of the expressions

$$\tfrac{1}{2}\rho w^2 \text{ or } \frac{\mu \mathrm{H}^2}{8\pi}.$$

Therefore, for the field excited by the single spherical moving charge e

$$\frac{w}{v} = \frac{e \sin\theta}{r^2} \sqrt{\left(\frac{\mu}{4\pi\rho}\right)}.$$

At the equator of the sphere, $r=a$ and $\theta=90°$, so that w will there attain its maximum value, say w_0, such that

$$\frac{w_0}{v}=\frac{e}{a^2}\sqrt{\left(\frac{\mu}{4\pi\rho}\right)},$$

and hence

$$\frac{w}{w_0}=\frac{a^2}{r^2}\sin\theta,$$

so that the ether flow in a tube of magnetic induction surrounding an electron, although very intense at its surface, becomes insensible at very small distances. Lodge suggests that w_0, being excited by v, can hardly have a higher value than v. Now, the smaller the value of w_0 the greater will be the value of ρ, the ether density; so, to obtain a minimum value for ρ, he assumes $w_0=v$. Then we shall have

$$w=\frac{a^2}{r^2}v\sin\theta,$$

from which it easily follows that the whole energy of the magnetic circulation due to the moving electron is

$$\frac{4\pi\rho a^3v^2}{3}\text{*}.$$

But if m is the mass of the moving electron, its total energy will also be represented by $\frac{1}{2}mv^2$, so that

$$m=\tfrac{8}{3}\pi\rho a^3,$$

which is twice the mass of ether contained within the sphere, so that the first, second, and third approximations to the value of m are in the ratio

$$1:\tfrac{3}{2}:2,$$

and they all represent the effective mass of an electron as due to a disturbance moving and identified with itself, affecting a region comparable with its volume. Now, the mass of an electron is known to be of the order 10^{-27} grammes, and its volume to be of the order 10^{-39} cubic centimetres, which gives for the density of the ether 10^{12} grammes per cubic centimetre, that is to say, about a billion times the density of water.

* For the total energy is

$$\int_0^\pi\int_a^\infty \tfrac{1}{2}\rho w^2 \,.\, 2\pi r\sin\theta \,.\, rd\theta \,.\, dr=\pi\rho a^4v^2\int_0^\pi\int_a^\infty\frac{dr}{r^2}\,.\,\sin^3\theta d\theta=\frac{4\pi\rho a^3v^2}{3}.$$

Taking $\rho = 4\pi\mu$ (see p. 32), the absolute value of μ for free ether will be about 10^{11} grammes per cubic centimetre.

Now,

$$K\mu = \frac{1}{c^2} = \frac{1}{10^{21}}.$$

So that the absolute value of the dielectric constant for free ether is 10^{-32} cubic centimetres per erg.

The electrostatic rigidity (Maxwell's coefficient of electrical elasticity) $\frac{4\pi}{K}$ is therefore 10^{33} dynes per square centimetre. The intrinsic constitutional kinetic energy of the ether will therefore be comparable with 10^{33} ergs per cubic centimetre, that is to say, to employ Lodge's illustration, with 100 foot-pounds per atomic volume. To employ another illustration given by Lodge, the total output of a million kilowatt electric power station for thirty million years, exists permanently, and at present inaccessibly, in every cubic millimetre of space.

Sir Oliver Lodge shows, in the concluding chapter of the recent edition of *Modern Views of Electricity*, that the value which he gives for the density of the ether leads to the result that "the amplitude of a wave of light, in a place where it is most intense, namely, near the sun, where its energy amounts to 2 ergs per cubic centimetre, comes out only about 10^{-17} of the wave-length. The maximum tangential stress called out by such a strain is of the order 10^{11} atmospheres."

Another interesting conclusion is that the velocity of the magnetic ether flow for a field of 12,000 C.G.S. units would only be about four inches an hour. Close to an electron it would attain a value comparable to its orbital speed, but at a distance of only a millimetre this would fall off to an infinitesimal amount, being less that a *millimicron* per century.[1]

The distinction drawn (p. 182) between the electromagnetic mass of an electron due to the ether twist arising from its motion, and the non-electromagnetic ethereal mass due to the twistless flow arising from the motion of the vacuole, or vacuous spherical hollow, constituting the nucleus, appears to me one of very great importance. Although the latter portion of the inertia appears to be negligible in comparison with the former in all electromagnetic phenomena, it is only through this comparatively minute effect that the non-electromagnetic forces requisite for the determination of the scale can arise, and it is to these non-electromagnetic actions that we must apparently look for the

[1] A *micron* being the thousandth part of a millimetre, a millimicron is the millionth of a millimetre.

explanation of gravitation and cohesion, as we shall see in Chapter XXIII.

The distinction appears to have been very generally overlooked, and the ether flow which we have identified with magnetic induction has been assumed to constitute in itself a magnetic field, even when unaccompanied by electric displacement. We have seen, however, that magnetism is not a fundamental conception in the electromagnetic scheme, being only a convenient method of treating statistically certain phenomena which cannot be dealt with by a consideration of the individual constituents, for the simple reason that our analytical methods are not sufficiently powerful.

Sir J. J. Thomson has shown, in the first chapter of *Recent Researches*, that the electromagnetic field may be completely specified in terms of Faraday tubes, that is to say, in terms of the electric displacement, without considering the tubes of magnetic induction at all. It follows from this that the representation of the field by means of the magnetic tubes of induction, or the lines of magnetic force, although convenient for many purposes, is merely a representation of certain statistical effects of electric displacement. The existence of a magnetic field involves, in fact, a variation, either in time or space, of electric displacement, and it is to the ether twist, and not to the accompanying twistless flow, that the electromagnetic action of a magnetic field is due. The twistless flow itself can affect the electrons only translationally, by carrying them along with it bodily. A twistless ether flow arising from other causes than electric displacement does not, and cannot possibly, involve the existence of a magnetic field.[1]

The manner in which ether displacement forms the necessary accompaniment of varying electric displacement is illustrated in fig. 2, p. 67, explanatory of the method of propagation of an electromagnetic wave, on the assumption that it is transmitted by bodily motions of the ether. This is the simplest and most natural

[1] I find from a note of Sir Oliver Lodge's (*Phil. Mag.*, June 1907), that Professor W. M. Hicks pointed this out in a footnote to his address to section A of the British Association in 1895, in which he observes:—"To prove this, consider a straight conductor moving parallel to itself and perpendicular to a uniform magnetic field. There exists a permanent potential difference between its ends. If, however, the field consists of a flow of ether, the effect is the same as if the conductor is at rest, and the direction of the magnetic field shifted through an angle. But this is the case of a conductor at rest in a field, and there is therefore no potential difference between the ends. Hence a magnetic field must consist of some structure across which a conductor cuts. A field may possibly demand a flow of the ether, but if so, it must carry in it some structure definitely oriented at each point to the direction of flow." See also C. V. Burton, *Phil. Mag.*, vol. xvii., 1909, p. 647.

assumption to make in accounting for electromagnetic action in terms of ethereal action. From the abstract point of view we may, in the interest of generality, bear in mind that the actual variations in the ether may be of a very different character, quite inconceivable to us. The assumption of actual bodily motion of the ether is, however, the most natural one, and indeed the only way in which we can form a clear mental picture of the actions taking place in an electromagnetic field.

In considering the phenomena of induced currents, and in many other cases, the problem is greatly simplified by tracing out the tubes of magnetic induction, and remembering that these tubes indicate a distribution of displacement such that a linear conductor crossing them in a direction at right angles to its length will, under the conditions stated on page 201, be acted on by an electric force in the direction of its length.

A representation of the phenomena of electrostatic induction may be obtained in terms of aggregates of simple vortex atoms. The presence of an electrically charged body in the ether will give rise to a change of twist in the surrounding ether. If a material conductor is now brought into this field of strain, the instability produced by the presence of the material atoms may be supposed to diminish the twist elasticity of the ether within it. The ether surrounding the conductor will then adjust itself to a new state of equilibrium, the lines of twist being altered so as to strike the conductor everywhere at right angles, and the conductor will be charged by induction, that is to say, there will be an alteration in the distribution of electricity over its surface. The conductor will not, however, receive an absolute charge, for the electric displacement is a stream vector, and therefore its flux into any closed surface, described entirely in the ether, so as to surround the conductor, will be null. Let us now suppose a thin filament of ether, connecting the two conductors, to have its twist elasticity diminished, say, for example, by the introduction of a thin connecting wire. The state of equilibrium will then be disturbed again, and the sudden loss of elasticity will cause a wave of alternating twist to roll backwards and forwards, along the surface of the filament, from one conductor to the other, and this oscillation will persist until damped by radiation or viscous action. As free ether has no viscosity, the damping in this case will be entirely due to radiation.

The vortex atom theory will therefore enable us to account for the communication of an electric charge to a conductor with the assistance of the supplementary hypothesis that a conductor is a region in which the elasticity of the ether is greatly reduced, the reduction being to zero for a perfect conductor, supposing such

a body existed. It might be possible to imagine vortices of such a character as might be expected to affect the etheric twist elasticity in this manner, but on the electron theory all that we have to assume is that the ions are free to move independently in *all* conductors, as we know experimentally to be the case in electrolytes, and in gases in the conducting state. The view propounded in this work, and which has already been suggested in Chapter IV., is a far more fundamental one.

It is a very remarkable fact, as Larmor observes, that although the gravitating matter of the universe is aggregated in sensible amounts only in excessively widely separated spots, yet, wherever it occurs, it appears to be composed of the same limited number of what are known as chemical elements. As was pointed out by Thomas Graham in his *Chemical and Physical Researches* in the earlier half of the last century, we are almost driven to explain this by supposing the atoms of all the chemical elements to be built up of combinations of a single type; and the considerations brought forward in the present chapter forcibly suggest that this single type is to be found in some such structural modification of the ether as is exemplified in Larmor's electric doublet.

If all material atoms were not composed of similar ultimate elements, it would not seem possible to account for the fact that the various chemical elements, in all their numerous combinations, are invariably associated with the same constant of gravitation. This fact, and the phenomena considered in Chapter VII., lead to the conclusion that all the sub-atoms must be quantitatively alike, and therefore we must assume that all electrons, pairs of which constitute our elementary doublets, are quantitatively equal with regard to the amount of ether strain pertaining to each, the only essential difference in electrons consisting in their being related in pairs as an optical image, or perversion, is related to its object, or as a right-handed screw is related to a left-handed one. On such a view electric transfer from ion to ion will arise, without any breaking down of the continuity of the ether, by the simple interchange of electrons by convection.

The most obvious supposition to make would be that the positive and negative electrons are both equally free to move in conductors, and therefore, that in all cases of transfer of electricity by interchange of electrons, they drift equally in opposite directions under the action of electric force. We shall see, however, that while electrolytic conduction is effected by the drifting in opposite directions, although not at equal speeds, of the oppositely charged ions, there are good grounds for believing that metallic conduction is effected by a comparatively small proportion of the

negative electrons, the positive electrons, and the greater proportion of the negative ones, remaining bound in the molecules of which they are constituents.

The theory developed in the last chapter and the present one shows that convection currents of electrons will give rise to the phenomena experimentally associated with electric currents, and that they lead to results in accordance with observation in such cases as are presented by the phenomena of magnetism, in which the theory of a continuous flow leads to discordant results. Apart from the electron theory, a convection current, due to the motion of a charged body, would in no way differ in its electrodynamic action from an ordinary conduction current, for when a charged body moves relatively to the surrounding ether, with a speed which is low in comparison with that of radiation, it practically carries its electric displacement system along with it as an equilibrium configuration. The electric displacement at any fixed point in the ether will therefore change continuously as long as the motion lasts, and the field will therefore be virtually filled with electric currents completed along the lines of motion of the charged elements of the moving body.

Rowland attempted, in 1878, to demonstrate experimentally the electrodynamic effect of convection currents by giving a high speed of rotation, in its own plane, to a glass disc coated with gold leaf, and he satisfied himself that a distinct, though small, effect was produced. The experiment was repeated by numerous physicists, some of whom failed to verify the effect, or attributed any effects observed to the induction currents, commonly known as Foucault currents, arising from the action of a magnetic field, due in this case to the earth's magnetism. In order to obviate this disturbing effect, the experiment was repeated in a different form by Rowland and Hutchinson in 1889. Two parallel glass discs were employed, rotating in the same direction, and in order to eliminate Foucault currents, the conducting surface was divided into segments by radial scratches. The results so obtained were confirmatory of the original experiments.

It will be interesting and instructive to consider the theory of this experiment, for, unless the charges are of the discrete nature involved in the electron theory, it does not appear that any effect should have been observed in the original experiment, although it should be obtainable in the later modification. If the gold leaf were a perfect conductor, there could not be any effect at all, on the earlier theory, when the surface was continuous. For the free oscillations on such a conductor are certainly independent of ohmic resistance, and it is difficult to imagine that they can be sensibly affected by any superficial viscosity; and it would there-

fore appear that, as far as regards the effect in question, the disc may be treated as a perfect conductor.

On the electron theory, on the other hand, such an effect would be produced, except in an absolutely perfect conductor, of which no example is known; for, although the phenomena of electric waves in conductors show that the electrons must be extremely mobile and sensitive to electric forces, this is in consequence of their strong charges relatively to their ordinary mass actions when aggregated into material bodies, and it does not by any means imply that, when the conductor is set in rotation, the electrons will slip backwards over its surface, owing to their electric inertia, and therefore fail to participate in the motion of the disc. There will be some backward slip when the rotation is started, but, after a time, which is in all probability an extremely short one, though long in comparison with the period of a vibration, the viscous actions to which the electric resistance of conductors is due will ultimately make them move with the rotating conductor.

When the disc is divided by scratches into separate portions an effect is to be expected on either theory, for each separate portion into which the conductor is divided will carry its field of electric displacement along with it, and this field will conserve its static configuration at any speed that is practically possible. In the later experiments, with two parallel discs divided up in this manner, the electric displacement will be across the field from one disc to the other, and will remain steady throughout the motion, so that the convection currents will be completely represented by the simple convection of the electric charges in the discs, instead of being distributed over the surrounding dielectric.

If an uncoated disc of glass or any other solid dielectric, carrying an electric charge, were set in rotation in a similar manner, it would carry its system of electric displacement along with it, causing convection currents, which would give rise to electrodynamic effects.

Röntgen, in 1888, detected the effects of convection currents produced by rotating a dielectric disc between the two plates of a charged condenser. In this case the displacement system, according to the unmodified Faraday-Maxwell theory, would maintain its configuration in space absolutely unchanged; so that, on this theory, no effect of the kind should be obtained, but should be obtained according to the electron theory. For, when a material dielectric is moved across an electric field, every constituent electron will produce its own convection current, consisting in part of change of electric displacement in the surrounding free ether, but completed, and converted into a stream, or made circuital, by the actual convection of the electronic charge itself.

When, as in Röntgen's experiment, the configuration in space remains constant during the motion, so that there is no change of electric displacement in the surrounding ether, the current is entirely due to the convection of the electrons, and this convection current then becomes circuital.

We see then that, unless the charge on a conductor is made up of discrete portions, separated by dielectric intervals, however small, as is the case according to the electron theory, the steady rotation round its axis of a charged conductor symmetrical about the axis, and therefore giving rise to an electric field which is also symmetrical about the axis, has no effect on the field of electric strain in the surrounding ether, and cannot therefore produce a magnetic field. There is another way in which we can see that this must be the case, for if a charged conductor merely involves a field of self-locked strain in the surrounding dielectric, the electric elasticity breaking down at the surface of the conductor, the rotation of the conductor about an axis of symmetry can exert no grip on the field of strain, and the latter will therefore remain unaffected. The convection of an isolated electric charge, on the other hand, carries with it its own electric field, and so alters the electric force at every point in the ether; the result of which is that a magnetic field is produced, the reaction of which tends to oppose the motion of the charge.

The consideration of this problem leads naturally to the cognate one of the electrification induced in a conductor by rotation in a symmetrical magnetic field, that is to say, the phenomenon known as Unipolar Induction, in which an electric current is produced by the rotation of a magnet symmetrical about an axis, about its own axis of symmetry, in its own field of force.

A simple method of exhibiting this phenomenon is shown diagrammatically in fig. 14, which represents a magnet pivoted so that it can be made to rotate about its axis NS by means of an endless cord passing over the pulley P which is rigidly fixed to the magnet. Two insulated metallic strips make rubbing contact at A and B with the magnet as it rotates, and are connected through a galvanometer. The magnet's lines of induction will evidently be symmetrical about the axis NS, so that the magnetic field would appear to be unaffected by the rotation. Nevertheless, a current will be observed to flow through the connecting wire, which will be proportional to the speed of rotation, and which will be reversed in direction when either the polarity of the magnet or its direction of rotation is reversed, just as if the lines of magnetic induction were physical entities rigidly attached to the magnet and rotating with it, in which case the rotation would cause a continual cutting of the lines of induction by the circuit

formed by the connecting wire and the portion of the magnet included between the sliding contacts.

Maxwell's equations of electric force, as modified by him to meet the case of a body moving through the ether, indicate an E.M.F. acting round the circuit formed by the connecting wire and the portion of the magnet included between the sliding contacts. It has been shown, however, by Larmor, that Maxwell's equations for moving bodies are incorrectly developed on the basis of the theory that the electrodynamic energy can be legitimately specified in terms of the elements of a continuous current and their mutual configurations, and that when

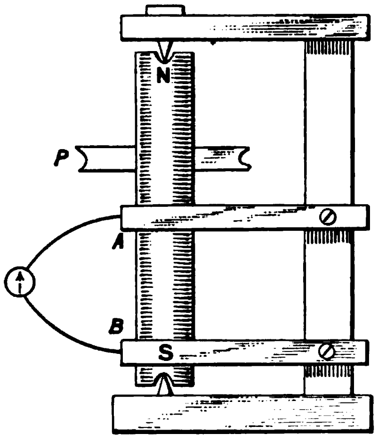


FIG. 14.

the requisite corrections are made, no electromotive force is indicated. He points out, moreover, that, on this theory, the energy belonging to a current element associated with an element of volume of the conductor would remain invariable, on account of the steady configuration of the motion, so that no electric force would arise in such an element, and that for the same reason there would be no electric force induced in the conductor. The electromotive force of the current actually observed to flow round the conductor would therefore have to be sought for wholly in the sliding contacts, which, considering the definiteness of this electromotive force and the great variety of types of contact that are possible, seems to be an untenable alternative.

The view which I endeavoured to maintain in a series of articles on Unipolar Induction which appeared in *The Electrician* in the

year 1888,[1] viz. that the lines of magnetic induction must be considered as rotating with the magnet, is shown by these considerations to be incorrect. In the case of a magnet unsymmetrical about the axis this would be the case, and I argued that, as there would be such a rotation, however small were the deviation from symmetry, it might be assumed, on the principle of continuity, to hold in the limiting case of absolute symmetry. It had not then occurred to any physicist to question the validity of Maxwell's equations, and this appeared to be the only possible way to account for the observed phenomena.

The electron theory affords a perfectly clear and simple explanation of the phenomenon. The passage of the electrons constituting the molecules of the magnet, through the stationary magnetic field established by the magnet in the ether within and around it, gives rise to an electric force acting upon each one of them, tending to produce electric separation, and therefore accompanied by a drift of electrons round the circuit.

Many more physical phenomena might be enumerated for which the electron theory accounts in a satisfactory manner, but which do not appear capable of explanation on the older theory. Sufficient evidence has, however, now been adduced to show that the electron theory is a necessary, as well as a legitimate, development of the Faraday-Maxwell theory, and we may therefore proceed to the explanation of further phenomena in their due order, on the basis of the electron theory alone, without investigating in every case whether or not the older theory would prove sufficient. The examples already considered, while they do not in any way invalidate the general principles of the Faraday-Maxwell theory, show conclusively that the distribution of electricity in an electric current is of a discrete nature, and that its representation as a continuous one leads to results which are in conflict with those derived from observation.

It may possibly occur to the reader as an objection to the view of the constitution of matter here set forth, that a rapidly revolving system of electrons would rapidly lose by radiation the kinetic energy essential to the maintenance of its existence. This does not, however, appear to be the case, for, as Larmor points out, it is a generally valid dynamical principle that any mechanical system whose motion is determined by imposed conditions of a steady character, will acquire a steady condition of motion after the preliminary oscillations, if any have been excited by some disturbance, have disappeared by radiation or viscosity. A system of electrons moving steadily across the ether, or revolving steadily round a centre, will therefore carry an unaltered configuration of

[1] *The Electrician*, vol. xxi. pp. 139, 171, 203, 233, 262, 300, 365.

strain with it, so that no loss of energy will occur except when this steady state of motion is disturbed. Under the influence of such disturbances radiation will take place, and unless this is compensated from sources external to the atom, the possibility of which is considered in later chapters, losses of energy must take place, which will ultimately lead to the breaking up of the atom, and the regrouping of its constituent doublets to form atoms of different type or types. The theory therefore receives additional confirmation in that it accounts for the observed phenomena considered in the chapters dealing with the subject of radioactivity.

In Larmor's formal development of the theory, after obtaining a system of equations for the propagation of electromagnetic disturbances in the free ether, he proceeds to extend the method of treatment to ether containing electrons, which, on the present theory, includes all material substances. The nucleus of an electron is assumed to be a very minute region, but not so small as to be practically indistinguishable from a mathematical point; and it is also assumed that there is no inertia or energy foreign to the ether residing in its nucleus such as would prevent free, unresisted mobility, so that nothing is involved in the electron except a strain form. It is difficult, as Larmor observes, to see how the second condition could be contravened. The former assumption is made, in the first place, to avoid the occurrence of indeterminate terms in the summations required in the mathematical analysis, but, at a later stage, physical reasons are adduced showing that it must necessarily be correct. If the electronic nuclei were devoid of sensible volume, the Newtonian principle of dynamical similarity would lead to the result that, given any existing steady system of electrons, the same system altered to any other scale of linear magnitude would be possible, supposing the only actions to be electric ones. There would, therefore, be no definiteness of scale in the molecules of material systems, which is not a condition which can well be contemplated as a possibility. J. H. Jeans has shown (see Appendix H), however, that this condition is not sufficient to ensure definiteness of scale, but that this result would be secured by the existence of non-electromagnetic forces, in addition to the electromagnetic ones, acting upon the electrons.

All that is requisite to the validity of the mathematical analysis is that the dimensions of the nuclei should be so small, compared with the distances between electrons, that these dimensions should not be sensibly involved in the forces between them. Strong confirmation of the validity of this assumption is afforded by the explanation given by the electron theory of the null result of the

Michelson-Morley experiment, referred to in Chapter VI., and which will be further considered in the present chapter; the change in dimensions of material bodies, due to motion through the ether, depends on the validity of the foregoing assumption, and would otherwise not be of the amount required to produce compensation.

Its validity is further confirmed by our experimental knowledge of the negative electron, the only kind as yet certainly known to us in the isolated state. The determinations of its mass and charge show that its diameter cannot well be greater than about 10^{-13} centimetre, or about a hundred-thousandth part of the diameter of an atom, which is of the order of 10^{-8} centimetre, that is to say, the ten-millionth of a millimetre. Sir Oliver Lodge observes that, on the basis of these measurements, if an electron were magnified to the size of the earth, the correspondingly magnified atom would occupy a sphere with the sun as centre, and four times the distance of the earth as radius. That is to say, the electrons in an atom must be about as far apart in that atom, in proportion to their size, as the planets in the solar system are in proportion to their size.

Sir J. J. Thomson, in his Presidential Address to the British Association at Winnipeg, states that recent experiments by Willisch at the Cavendish Laboratory appear to afford experimental evidence of the existence of positive electrons. When mixtures of methyl iodide and hydrogen in varying proportions were exposed to Röntgen radiation, the positive and negative particles set free were found to originate almost entirely from the methyl iodide, and nevertheless they were observed to travel with the same velocities as when originating and travelling in pure hydrogen. Similar results were obtained when carbon tetrachloride and mercury methyl, respectively, were substituted for methyl iodide. The mass of a positive particle was found to be comparable with that of an atom of hydrogen, which made it difficult to be quite certain of their isolation, as there was the possibility that the true positive electrons might consist of much smaller bodies remaining attached to the hydrogen atoms.

The comparison of an atom with a solar system affords a very real assistance in forming an adequate conception of the nature of material bodies in accordance with our present point of view. Subsequent chapters will show, moreover, that the cycles of changes in material atoms, as observed in the case of the more unstable forms, such as the constituents of the substances specially designated as radioactive, correspond so closely with those observed in star systems that no absurdity or impossibility would be involved in a speculation which regarded the whole starry system of our visible universe as forming nothing more than an

infinitesimal portion of the substance of some greater universe. It might even form but a single molecule in some particular material body of such a universe. The matter composing such a universe would be far more complex than any form of matter known to us, but its physical properties might quite possibly bear at least a general resemblance to those of the matter with which we are acquainted. To seriously elaborate speculations on these lines would be unprofitable, for they would have to remain speculations, and could neither be proved nor disproved by any means of investigation conceivable to us. As an illustration, however, it is not unprofitable, for it enables us to grasp the fact that if matter is constituted in accordance with the theory we are developing, a particle of any ordinary material body, if magnified to such an extent as to fill the space occupied by the whole visible universe, would resemble, in its general character and properties, the complex of star systems of which that universe is composed. It is, moreover, of material assistance in enabling us to grasp the possibility of solid material bodies moving freely through ether of a density of which we can hardly hope to form any adequate conception. It shows us that, as Sir J. J. Thomson expressed it in the Presidential Address previously referred to, what we call solid matter really consists mainly of holes. An approach to a mental representation of the passage of such a solid body through the ether may be obtained by imagining a very open network of the finest steel wire to be drawn through a mass of mercury or melted lead. The analogy is, however, extremely inadequate, since the progress of the network would be retarded by the viscous, or frictional, resistance of the mercury or lead, whereas the ether, being a perfect fluid, offers no frictional resistance, its retarding effect being equivalent merely to a slight increase in the diameter, and consequently in the mass, or inertia, of the wires forming the network, which, except for this, must be pictured as moving through the mercury or lead as freely as through a theoretically empty space.

On the basis of the foregoing assumptions, the fundamental equations of equilibrium and motion of free ether will still be sufficient to trace the natural sequence of change in the complex medium composed of the ether and its contained electrons, considered as mere strain forms. Such an analysis would, however, be of no assistance in studying the electrodynamic phenomena of material bodies, in which we can only experiment upon, and observe the behaviour of, matter in bulk. We cannot take any direct cognisance of the separate molecules, much less of the separate electrons of which they are formed. The mechanical theory must therefore deal with elements of volume in which all

that we can in general take direct cognisance of will be the excess of either positive or negative electrons, constituting a volume density of electrification, and the movements of electrons in such an element of volume must be dealt with in the form of the aggregates which constitute fluxes and molecular whirls. The molecular theory, however, still retains its value in enabling us to form and test various hypotheses of molecular structure and arrangement, with a view of accounting for the distinctive features of the mechanical phenomena actually observed. Now, when we cease to consider merely the ether between the molecules, and turn our attention to a material medium, in which the irregularities due to the discrete electrons are considered as smoothed out into a continuous distribution, the specification of electrical action directly in terms of the ether becomes inapplicable.

In order to ascertain what is to take its place, we will consider the effect of moving an electron $+e$ from a point P to a neighbouring point Q. This will give rise to a strain flux from Q to P along lines beginning at P and ending at Q, the addition to the initial distribution of strain in the medium being clearly the electric displacement due to the formation of the doublet $-e$ at P and $+e$ at Q. This additional flux from Q to P will not in itself constitute a stream, but the flux circuit may be completed by the addition of a linear flux of the same total amount e from P to Q, entirely along the one line PQ. If we complete in this manner the fluxes of *ethereal electric displacement* arising from the changes in position of all the electrons in a unit element of volume by the fluxes of these *true electric charges* through the ether, we shall obtain a stream, or circuital, flux which may be considered as the total electric displacement per unit volume. In a material dielectric this true flux of electrons will represent the polarisation per unit volume, and will give rise to an electric moment per unit volume exactly analogous to magnetic moment per unit volume, and which may be defined in a similar manner. In a circuit consisting entirely of conducting material, the electrons will circulate completely round the circuit, forming a current of conduction, and, except in the case of oscillations of a frequency approaching those concerned in radiation, the ethereal electric displacement will be of negligible amount in comparison with the *true current* consisting of the electron circulation. We are now in a position to grasp all the circumstances of the current flowing round the simple condenser circuit illustrated on page 167. Suppose, as before, the plate A, fig. 15, to become positively charged, and B negatively. When the plates are at rest relatively to each other, the electric displacement will be everywhere perpendicular to the conductors; but when approach takes place, the compen-

sating flux along the wire due, according to our present point of view, to the motion of its electrons, will cause the electric displacement in the dielectric in the immediate neighbourhood of the wire to be directed along the surface of the latter if it is a perfect conductor, and very nearly so in the case of a metallic wire. That is to say, an electric force will act along the connecting wire in the direction ACB, giving rise to a flow of electrons round the wire and a corresponding electron flux across the dielectric, as indicated by the arrows. The electric force across the dielectric separating the plates may be considered as due to a right-handed twist of every volume element of the ether about an axis perpendicular to the plates, that is to say, parallel to, and in the direction of, AB. The lines of twist, viz., the lines representing the direction of the axis of twist at every point, which emerge from the lower half of the plate A, after traversing the

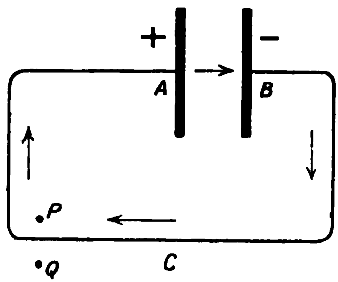


FIG. 15.

separating dielectric, will be deflected towards the connecting wire, owing to its forming a channel of greatly reduced elasticity through the ether, and will pass round, more or less parallel to the wire, according to their nearness to it, until they enter the lower half of the plate B, and so complete their circuits. The lines of twist emerging from the upper half of the plate A will, for the most part, complete their circuits in a similar direction outside the connecting wire, but some of them will do so above the plates. We may define a right-handed strain form, or positive electron, as that terminal of a Faraday tube in which the ether twist in the tube is clockwise to an observer looking from the positive towards the negative end. Looking in the same direction, the twist at the negative end will be counter-clockwise, and this terminal will constitute a left-handed strain form. As the direction ACB round the circuit is that of decreasing electric force, that is to say, of decreasing right-handed twist, the positive electrons, being right-handed strain forms, would diminish the

strain by moving along the connecting wire in this direction, and therefore, if free to move, which, as we have seen, is probably not the case, this would be the direction of their motion. The negative electrons, on the other hand, being strain forms with left-handed twist, would diminish the ether strain by moving in the direction of increasing right-handed strain, that is to say, in the direction BCA. If the wire were a perfect conductor, assuming such a thing to be possible, there would be no ether elasticity at all within it, and the effect of the ether twist would therefore be entirely communicated to the internal electrons at the surface of the wire.

The relaxation of ether strain within the conducting wire will spread outwards from the wire into the surrounding field, the diminution of resistance to twist becoming less as the distance from the wire increases. The twist spreading outwards from the space between the condenser plates towards the wire will therefore, as long as the current remains constant, form a steady configuration of increasing twist as the distance from the wire becomes less. As in the case of the wave transmission considered on page 66, this rotation of each element of volume of the ether is equivalent to an ether displacement or flow. The flow due to the twist of a single volume element would evidently take place round its circumference, and be in the plane perpendicular to the axis of twist. Uniform twist about parallel axes throughout any region would give rise to flow round the boundary, the flow at every point being perpendicular to the direction of the axes of twist. In the neighbourhood of a straight wire the axis of twist is sensibly parallel to the wire; the twist is of the same amount for every volume element having its centre on a circular cylinder having the wire as its axis, and the amount decreases as the diameter of the cylinder increases; that is to say, it decreases as the distance of the volume element from the wire increases. The tubes of ether flow, viz. of magnetic induction, are therefore contained between concentric circular cylinders surrounding the wire. The lines of magnetic induction in a plane section of the field intersected by the wire at a point O are therefore circles with O as centre. When the approach of the plates ceases, the electric displacement will rapidly resume its distribution of equilibrium, in which it is everywhere perpendicular to the conductors, and the magnetic field will then disappear.

In this example we have a wave of electric displacement approaching the conducting wire from all directions, the direction of the displacement near the wire being sensibly parallel to it. The Faraday tubes near the wire are therefore parallel to it and approaching it. Applying, therefore, the diagram of fig. 13, p. 179,

to a point P near the wire and in the inside of the circuit, the direction of the magnetic induction will be perpendicular to the plane of the diagram and upwards; if, on the other hand, we apply it to a point Q just outside the circuit, the magnetic induction will be vertically downwards from the plane of the diagram. These are the same directions as are given by Ampère's rule for the motion of a magnetic north pole in the neighbourhood of a current in the direction ACB. Since the ether flow round the tubes of magnetic induction is twistless, it is unresisted, and cannot therefore in itself give rise to electrodynamic actions. The tubes of magnetic induction must not, therefore, be regarded as directly operative, but only as being inseparably connected with the motion of the Faraday tubes of electric displacement at right angles to themselves. Whenever, for example, a linear conductor cuts across tubes of magnetic induction at right angles,

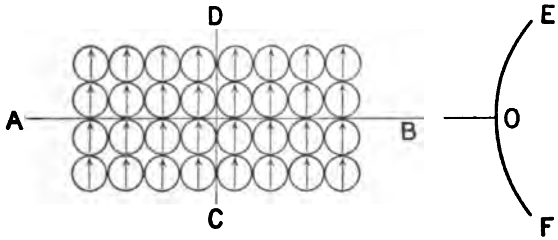


Fig. 16.

unless its direction of motion be at right angles to the gradient of the electric force, that is to say, to the gradient of ether twist,[1] the conductor will be traversed, during its passage across the tubes, by Faraday tubes having a component in the direction of the length of the conductor, and moving at right angles to their own length. There will therefore be electric displacement along the conductor, and consequently an electric current if the conductor forms a closed circuit, lasting as long as the passage continues.

Let us now attempt, by means of the Kelvin model of the ether, to trace out in detail the actions which take place in the electromagnetic field when a straight linear conductor of circular section, forming part of a closed circuit, is made to approach a fixed parallel straight linear conductor, also of circular section, traversed by a constant current.

Let fig. 16 represent a greatly magnified section of the field by a plane perpendicular to the fixed conductor, when there is no

[1] The limitation expressed by the sentence in italics should be specially noted, and is considered further on page 204.

current in the latter. Let EOF represent a greatly magnified portion of the section of the fixed conductor, having its centre on the straight line AB produced beyond B, and at so great a distance from B, indicated in the diagram by the break in the line BO, that when a current is made to flow upwards perpendicularly to the plane of the paper through the wire of which FOE is part of the section, the small portion ACBD, of the magnetic field, may be treated as uniform. Let the section of each sphere representing a volume-element of the ether be marked by an arrow, perpendicular to AB, or parallel to CD, to indicate its position of equilibrium.

Now suppose a constant upward current perpendicular to the plane of the paper to flow through the conductor EOF. This will give rise to a distribution of twist of the character indicated in fig. 17. The electric force acting along the conductor EOF,

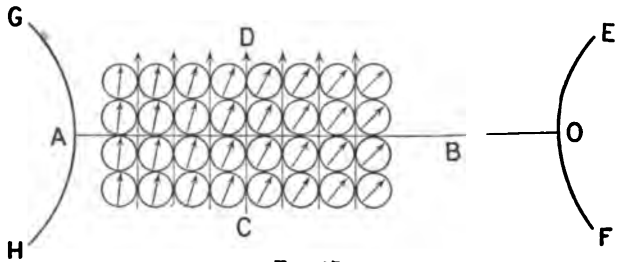


Fig. 17.

perpendicularly upwards with reference to the plane of the paper, will tend to move the positive electrons upwards, and will actually cause the negative electrons to move downwards. The resulting elongation of the Faraday tubes, which were of molecular dimensions before the passage of the current, will relax the clockwise, or right-handed strain within the tubes, looking downwards along the conductor, and therefore give rise to a right-handed twist in the surrounding ether.

Let us first consider simply the motions of the Faraday tubes. The elongation of immense numbers of molecular Faraday tubes within the conductor will force a large number out of the conductor into the surrounding ether. These will all be perpendicular to the plane of the paper, and will travel away from the conductor in directions radiating from its centre. In the small portion ACBD of the field, at a great distance from the conductor, these radii will be sensibly parallel to the central radius AB, so that the tubes of magnetic induction will be sensibly parallel to CD. That is to say, ACBD may be treated as a portion of a

uniform magnetic field, when the steady state is attained. Now let a second conductor, of which GAH is a portion of the section, this conductor being perpendicular to the plane of the paper, and forming part of a closed circuit, be made to move in the direction AB. As the tubes enter the conductor they will again shrink to molecular dimensions, causing a flow of electrons, and therefore a current, in the opposite direction to that in the primary conductor EOF.

Let us now trace the actions in the field directly in terms of the ether twist. The right-handed twist arising in the ether surrounding the tubes within the conductor EOF will spread into the ether outside the conductor, the amount of twist gradually decreasing as the distance from the conductor increases. The twist, being proportional to the electric force at each point, will be inversely proportional to the distance from the centre of the conductor, so that within the area ACBD, where the radii may be considered as parallel to AB, the twist will decrease uniformly from B to A, the decrease of twist in passing from any vertical column of volume elements to the column on the immediate left-hand side of it, being the same throughout. It follows, therefore, that the ether flow along lines parallel to CD, as indicated by the arrows, will be of the same value throughout the portion ACBD of the field, which may be regarded as uniform, which is as it should be, as it has been pointed out that the ether flow arising from the electric twist is proportional to the magnetic induction. Since the ether is a perfectly frictionless fluid, the circulation so established round the tubes of magnetic induction will continue unchanged as long as the distribution of twist is unchanged, that is to say, as long as the current is maintained constant in the primary conductor EOF. When the conductor GAH is moved in the direction AB, the twisted volume elements of ether which successively enter into the conductor will return to their positions of equilibrium, owing to the breaking down of the elasticity of the ether within the conductor, due to the mobility of the negative electrons, giving rise to a left-handed ether twist within the conductor, accompanied by a downward flow of current, that is to say, an upward flow of negative electrons. As long as the conductor continues to move uniformly in the direction AB, the steadily increasing twist of the volume elements entering the conductor will maintain a uniform current in it, but this will cease when it is brought to rest.

This example will be sufficient to enable the reader to trace out for himself the actions which take place when the current in the primary conductor is stopped or reversed, the secondary remaining stationary, and other simple cases. The observed relation

between the directions of the magnetic induction, the direction of motion, and the resulting current, or electric displacement, as indicated in fig. 13, establishes the correctness of the identification on page 199 of a right-handed strain-form, as there defined, with a positive electron.

It has hitherto been usual to state that electric force is always set up in a linear conductor when it cuts across the lines or tubes of magnetic induction, without the limitation to which attention is called in the footnote on page 201. Since the usual statement does not lead to results in conflict with observation, this point requires further consideration. The case excluded by the limitation which I have introduced would be exemplified by the motion vertically upwards or downwards of a wire lying along the straight line AB in fig. 17. It is obvious from the figure that such a motion could not give rise to any electric force along the conductor if the representation here given of the action of the field is a correct one. The linear conductor AB is perpendicular to the primary conductor, of which EOF is a partial section, and its direction of motion is parallel to the length of the latter. We know that under these circumstances no current will be produced in AB, but this has usually been attributed to the fact that every tube of magnetic induction is cut twice by the moving conductor, so that the resulting electric forces, being equal and opposite, cancel each other's effects. According to the representation here given, this explanation of the null effect is incorrect, the true explanation of it being that no electric force is produced at either point of section. An examination of the various possible cases shows that similar considerations apply to all of them.

An interesting case to consider is that of a longitudinally magnetised iron ring alternately linked and unlinked with a conducting metal ring. If both rings are continuous it will be impossible to perform the linking or unlinking. A gap must therefore be made in one of them through which the other can be pulled, and this gap may be occupied by any convenient liquid or gas, but not by a solid. We will also suppose that a separate gap is made in the ring forming the induced circuit, and that the ends are connected to the terminals of a galvanometer adapted to indicate the presence of an induced current.

If a second gap is now made in the conducting ring, and this gap is placed below the surface of mercury contained in a sufficiently large trough, the ring may be linked and unlinked with the magnetised ring without any current being induced in it, in spite of the fact that the magnetic flux circulating round the magnetised ring is made to cut across the conductor forming

the electric circuit, increasing the magnetic flux through the electric circuit by linking, and decreasing it by unlinking.

This is a case in which Faraday's general law of electromagnetic induction is inapplicable, because the magnetic system of the continuous ring is completely isolated, giving rise to no external magnetic effects, and conversely, being unaffected by external magnetic action. The surface of the magnetised ring therefore forms a boundary impervious to magnetic action, which divides the portion of the electromagnetic field dealt with into two portions, the portion contained within the ring and the portion external to it.

In order to get rid of this *cyclosis*, as it is called, the boundary must be cut across, making a gap in the magnetised ring, by means of which a connection is established, through the ether, between the magnetic system of the ring and the remainder of the electromagnetic field. When any portion of the conductor forming the electric circuit is made to pass across this gap, either inwards or outwards, an induced current will be observed.

If, after cutting the gap in the magnetised ring, the linking and unlinking were performed by pulling a part of the ring distant from the gap through the liquid part of the electric circuit, there would again be no induced current; for there would still be cyclosis in the portion of the electromagnetic field dealt with, that is to say, in the portion concerned in the mutual action of the two circuits.

These experiments have been suggested by one shown by Mr Carl Hering to the American Institute of Electrical Engineers in February 1908. A loop formed of two flexible strips, with the ends pressed together, had its circuit completed through a galvanometer. When the loop was passed over or drawn off the leg of a horse-shoe magnet in the ordinary manner, an induction current was of course observed in the galvanometer, but not so when the two circuits were unlinked by pulling the magnet out of the loop through the joint formed by the flexible strips. The object of Mr Hering's paper was to press the advisability of explicitly stating, in the ordinary elementary text-books, the conditions under which the Faraday-Maxwell law of electromagnetic induction is applicable. These conditions are made perfectly clear in Maxwell's mathematical treatment of the subject, but are not stated in a form that appeals to the understanding of the elementary student or of the compilers of many of the elementary text-books. I have attempted on page 16 to express the conditions in simple physical terms divested of mathematical difficulties. In the discussion on Mr Hering's experiment the existence of cyclosis in the field was naturally referred to as accounting for the failure of the general law

to express the facts. It does not, however, tell us what the facts are, only that we must find some other means of arriving at them.

From our present point of view the conditions in this case are extremely simple. The twistless ether circulation, due to the circulatory motions of the magnetic vortices arising from the rapid orbital motions of the electrons, and representing the magnetic induction, is here confined entirely within the substance of the ring, and flows round within it. Now we have seen that this ether flow, known to us as magnetic induction, is the necessary consequence of the elementary twists constituting the electric displacement, and that the direction of the electric displacement, that is to say, the direction of the axis of twist, is everywhere at right angles to the resulting ether flow. Supposing the ring to be a circular one, formed of circular wire, the electric displacement will be everywhere tangential to the circular sections formed by planes passing through the centre of the ring and perpendicular to its plane; that is to say, the electric displacement is everywhere in the average direction of the axes of the orbital motions of the electrons forming the magnetic molecules orientated, as indicated by the magnetic induction, so that the average orbital motion is as though all the electrons were revolving in the same direction round the circular sections of the ring. There is, therefore, no ether strain external to the ring, and no communication between the internal and external ether. The internal mass of molecular currents arising from these orbital motions will counteract one another's effects and be equivalent to a flow round the surface of the ring along the circular sections like the circulation in a smoke ring. When the electric circuit crosses through the gap in the magnetic ring, the direction of the resulting electric displacement, being perpendicular both to the magnetic induction and to the direction of motion, must be along the length of the conductor, so that a current will flow through the latter as long as its motion continues across the magnetic flux which is crossing the gap.

The foregoing considerations show that in all cases of electromagnetic induction the physical effect is due to electric displacement, acting at right angles to the direction of motion and to the magnetic induction, being produced within the material of the conductor in such a direction as to give rise to a current. In a linear circuit this must evidently be along the length of the circuit. The actual physical effects are therefore local ones produced in the material of the conductor while actually passing across the tubes of magnetic induction, and the law expressed in terms of the rate of decrease of the tubes of induction passing through the circuit is merely the expression of the result of summing up all these local actions round the circuit.

The expression of the law in terms of the area is very much simpler than it would be in terms of the cutting of the tubes by the conductor. The law in terms of the area of the circuit, as already given, is a general one applicable to all circuits linear or other. The law in terms of the cutting of tubes of induction may be simply expressed in the case of linear circuits, but is not capable of very simple expression in the general case. There does not, moreover, appear to be any advantage to compensate for this increased complexity, for the limitation to an electromagnetic field unaffected by cyclosis is inherent in the nature of the reasoning, and must therefore be retained in either form of expression.

Returning to the general development of the theory, we saw on page 198 that, in ordinary electrodynamic phenomena relating to currents of conduction, the electron flux is of far more importance than the averaged disturbance of the ether. Larmor accordingly proceeds to transform the general expression for the kinetic energy of the medium in terms of the magnetic induction into a form which expresses it as the effect of motion of the electrons, and in which the magnetism is implicitly included as molecular current whirls. A system of equations, similar to Maxwell's, for determining the electric force is then obtained for the case of a medium at rest.

The next step in the analytical development is to investigate the effect produced by the motion of a material system through the ether, by obtaining a set of equations expressing the state of rest or motion of the medium referred to an ideal framework moving with the material system.

The principal general result is that, whatever be the extraneous or imposed magnetic field, the distribution of charges on the system of conductors moving through the ether with a velocity v is identical with that of equal charges on a stationary system which is the same as the actual one uniformly elongated in the ratio of $1:1+\frac{1}{2}\frac{v^2}{c^2}$ in the direction of motion, c representing, as before, the constant of radiation. Now we shall presently see that there are good reasons for supposing that one effect of the motion is to actually cause a material system to shrink in this direction in the ratio $1:1-\frac{1}{2}\frac{v^2}{c^2}$. Neglecting $\left(\frac{v}{c}\right)^4$, the shrinkage will counterbalance the hypothetical elongation, and we arrive at the result that—

When a material system moves with the steady uniform velocity v through the ether, it shrinks in the ratio $1:1-\frac{1}{2}\frac{v^2}{c^2}$ in the

direction of motion, while the electric distribution throughout it, and the distribution of electric force around it, remain the same as if it were at rest.

If the moving system contains permanent magnets, the analysis shows that their motion through the ether would give rise to changes in the electric force, but all permanent magnets that we are acquainted with are conductors of electricity, and the analysis shows that the changes in the electric force would then give rise to surface charges on the magnets themselves of such amounts as exactly to compensate the external effects of the changes in electric force. Theoretically such effects might be produced by permanent magnets formed of dielectric substances, provided they were not electrostatically screened off from the moving charges by the interposition of conducting matter. No practical problem would, however, arise of this kind, as sensible permanent magnetism has never been observed in any dielectric.

These results are obtained by a comparison, developed so far as to be correct up to the first order of the small quantity v/c, of the ethereal vectors referred to fixed and moving frameworks respectively.

Up to this order the analysis confirms Maxwell's result in that it leads to an effect on the velocity of transmission of electromagnetic waves in agreement with Fresnel's formula. We have not yet, however, established the shrinkage of a material system in the ratio of $1 : 1 - \frac{v^2}{c^2}$ in the direction of its motion through the ether, suggested by Fitzgerald and Lorentz as a possible explanation of the null result of the Michelson-Morley interference experiment. This is a second order effect, and therefore needs a development of the analysis to an extent which will give results reliable up to the second order of the small quantity v/c, that is to say, results in which terms containing the square of v/c, as well as those containing only its first power, are taken into account. By means of a more complete analytical transformation Larmor has succeeded in effecting this further development, and arrives at the following result, which is correct to the second order:—

If the internal forces of a material system arise wholly from electrodynamic actions between the systems of electrons which constitute the atoms, then an effect of imparting to a steady material system a uniform velocity of translation is to produce a uniform contraction of the system in the direction of motion of an amount equal to $1 - \frac{1}{2}\frac{v^2}{c^2}$. The electrons will occupy corresponding positions in this contracted system, but the ethereal

displacements in the space around them will not correspond; that is to say, the electric and magnetic displacements at corresponding points in the two systems will differ in value. Since the electric and magnetic vectors of radiation both lie in the wave-front, the relative wave-fronts in the two systems correspond, as also do the rays which represent the paths of the radiant energy relative to the systems. The change of the time variable, in the mathematical transformation employed for the comparison of radiations in the fixed and moving systems, involves the Doppler effect on the wave-length.

When a train of radiation is travelling with a velocity V along the direction in which the material medium is itself moving with velocity v, the velocity of the train relative to the moving medium is found to be

$$\frac{V\left(1-\frac{v^2}{c^2}\right)}{1+\frac{Vv}{c^2}} = V - \frac{v}{n^2} - \left(\frac{1}{n} - \frac{1}{n^3}\right)\frac{v^2}{c}$$

to terms of the second order, where c/n is written for V, so that n is the refractive index of the medium. The second term in this expression represents the Fresnel effect, and the remaining term represents the second order correction which accounts for the Michelson-Morley negative result.

The two ethereal vectors, the electric displacement and the magnetic induction, are necessarily in the same phase in the free ether, since the former is then always proportional to the *curl* of the latter. They will therefore vanish simultaneously, so that regions of no disturbance in the fixed system correspond to regions of no disturbance in the moving system. Since optical measurements are usually made by the null method of adjusting the apparatus so that the disturbance vanishes, there must be a general absence of effect of the earth's motion in optical experiments, up to the second order of small quantities.

It will be instructive to consider the application of these results to the simple case of a pair of electrons of opposite signs, one of which is describing a steady circular orbit round the other, when the latter is at rest in the ether. Since such a system would form an effective vibrator, the orbital velocity must be supposed so small in this case that radiation is unimportant. The foregoing considerations then lead to the conclusion that when this pair is moving through the ether with a velocity v in a direction lying in the plane of the orbit, the orbit relative to the translatory motion will be flattened along the direction of v to the ellipticity $1-\frac{1}{2}\frac{v^2}{c^2}$, and that there will be a first

order retardation of phase in the orbital motion when the electron is in front of the mean position, combined with acceleration when behind it, so that on the whole, the period will be changed only in the second order ratio $1+\frac{1}{2}\frac{v^2}{c^2}$. The modification in the more general case of motion in a direction inclined to the plane of the orbit will be of the same character, and hence the results may be extended to include any molecule constituted entirely of electrons in orbital motion.

The fact that the free vibration periods of an atom are not affected to the first order by its uniform motion through the ether is of course vital to the theory of the spectroscopic determination of stellar velocities in the line of sight.

From the fact that the uniform motion of an atom through the ether does not disturb its constitution to the first order, nor the ethereal symmetry of the moving system fore and aft, we can deduce the important conclusion that when steady motion is established the mean kinetic energy of a material system consists, to the first order, of the internal energy of the atom, which is the same as when at rest, together with the sum of the energies belonging to the motions of translation of its separate electrons; for the total disturbance in the ether is obtained by the addition to the disturbances due to the internal motions of the electrons in the molecule, of those due to their common velocity of translation. Now, the total kinetic energy of the system is the value of the volume summation of the square of the ethereal disturbance, and it follows that this will be obtained by adding together the results of two summations, one being the sum of the squares of the disturbances, and the other the sum of their products taken in pairs. One factor in each product, viz. the one due to the uniform translation, is constant in time, and symmetrical fore and aft as regards each electron; while the other factor, due to the orbital motions of the electrons, is oscillatory, and symmetrical to the front and rear of each orbit: the total result of this summation is therefore null. But the sum of the squares of the disturbances represents the original internal energy of the electrons added to their energy of translation. The coefficient of half the square of the velocity of translation in the latter part is therefore, up to the first order, the measure of the inertia, or mass, of the atom.

Hence, to an order in which the square of v/c is neglected, the electric inertia of the system is the sum of the inertiæ of the electrons composing it. The law of constancy of mass throughout molecular transformations is therefore a consequence of our present view of the constitution of matter, always subject to the

condition that the velocities of the electrons remain sufficiently small for the square of their ratio to the velocity of radiation to be negligible.

It would at first sight appear that the correlations here considered between material systems at rest and in motion through the ether could only apply strictly to systems in which the constituent electrons were at rest relatively to one another. Let us first consider the case of dielectrics at the absolute zero of temperature, so that the molecules may be taken as fixed, though the electrons are in rapid orbital motion within the molecules. The only difference arising from this cause, however, will be the slight relative phase changes of the kind considered above, and it can be shown that these relative phase changes will have no effect upon the permanent or secular relations between the molecules, which are supposed to be sufficiently far apart not to interfere in a structural manner with each other so as to form compound molecules. The principle on which the relative phase effects are eliminated was first employed by Gauss in dealing with the perturbations of a planetary orbit by the disturbances due to the motions of other planets. Instead of proceeding by addition of the elementary effects produced by a planet as it moves from point to point of its orbit, Gauss showed that the secular results, as distinguished from mere periodic alternations, are the same as if the mass of the planet were supposed to be permanently distributed round its orbit in such a manner that the density at any point is inversely proportional to the velocity which the planet would have at that point. Just in the same way, the steady effects of the molecules, as distinguished from mere vibration effects, is the same as if the mass of each electron were distributed round its circular orbit, thus forming effectively a vortex ring, of which, however, the intensity is subject to variation owing to the action of other systems. The same principle enables us to eliminate the vibrational effects of the molecules of actually existing material systems.

The case is different, however, when there are electric currents flowing in the system, for that involves the transfer of electrons between different portions of the system which may be at finite distances apart. Now, if the current in conductors, such as metals, is carried by comparatively few electrons, they will travel with sensible speed, but the smallness of the number, as compared with the number of combined electrons, will prevent their changes of position from sensibly affecting the molecular structure of the medium. If a considerable proportion of the electrons take part in carrying the current, the same result will be attained by their extremely minute velocities. We know, indeed, as an experimental

fact, that the mechanical structure of a conductor is not sensibly affected when it carries a current. Some knowledge of the nature and amount of the effect of the earth's motion on electric conductivity can be obtained from the phenomena of electrolytic conduction (see p. 252).

Since magnetic induction is represented by ether flow, the fact that the magnetic permeability of all sensibly non-magnetic substances is almost identical with that of free ether, that is to say, of a vacuum, must be considered as indicating that the ether flows with practically its full velocity in all such media, so that very little obstruction can be caused by the matter, and therefore, in the motion of a body through the ether, the outside ether must remain at rest instead of flowing round its sides. This is exactly what we should anticipate from the nature attributed to matter according to our present point of view.

Experiments made by Sir Oliver Lodge in 1893, with a view of attempting to determine the convective effect of magnetic fields on the velocity of propagation of light, gave entirely negative results. The magnetic force employed was 1400 C.G.S., and the delicacy of the method was such that the negative result showed that the velocity of light cannot be altered by as much as two millimetres per second for each C.G.S. unit of magnetic force.

In the description on page 71 of the material model imagined by Lord Kelvin as an aid towards the mental representation of the properties of a fluid medium having elastic resistance to twist, it was pointed out that such a model can be made to represent a field of steady electric force lasting, without sensible decay, for any assigned length of time, although *not for ever*. The question whether a limitation of the same kind may not also be involved in the general scheme of a rotationally elastic ether is clearly one of the most fundamental importance. In developing the dynamical equations of the medium, it has been assumed that the effect arising from the velocity of convection of the elastic strain due to the motion of the strain forms through the ether may be neglected. Any such effects, which would be of the second order of small quantities, would be utterly insignificant in ordinary electrodynamics, and would be beyond the limits of optical observation. Nevertheless, as was pointed out to Larmor by Lord Rayleigh, in the case of a steady magnetic field maintained *for an unlimited time* by a current of electrons constrained to flow permanently round a circuit, the infinitesimal displacements of the elements of

volume of the ether arising from this cause might conceivably accumulate until their effects became sensible. If there were no other way out of this difficulty, it might be evaded by the assumption that there is no such thing in nature as an *absolutely* permanent magnet. Larmor shows, however, that a much more satisfactory solution is indicated by a further consideration of the nature of the elastic twist of the medium. When the change of position of an element of volume is of finite amount, it can no longer be definitely analysed into rotations and pure strains, or, in more generally familiar language, into twists and stretches, in such a way that the order in which these are effected may be indifferent. The ethereal elastic twist therefore, in such cases, needs more precise formulation. The nature of the necessary modification is indicated by Lord Kelvin's model, for in that model a relative stretch of the element of volume does not tend to twist the sub-element which is elastically effective; the efficient elastic twist is, in fact, the vector sum of the whole series of relative twists which the element of the medium has experienced in its previous history.

This more precise definition of the total twist, applied to the ether, makes the twist, from which the electric forces arise, equal to half the curl of the linear ether displacement, only when these quantities are so small that their squares and products are negligible, which will not be the case after the long-continued cumulative action of a permanent magnetic field. In that case, however, the small twistless velocity of flow, corresponding to the magnetic field, will contribute to the twist only by shifting the element of the medium along with the otherwise produced twist; and the twist so transferred will continually adjust itself by elastic action into the new equilibrium configuration. Unless the velocity of the ether is absolutely uniform, its movement by the steady magnetic field will, in consequence, lead to space variations in the twist, that is to say, to the development of electric charges throughout the volume. This development will be extremely slow, and, moreover, the combination of permanent electric and magnetic fields from which it arises must be confined to limited regions, outside of which the ether is in equilibrium. The electrification so developed must therefore be positive and negative in equal amounts, and these diffuse charges of the second order of small quantities will ultimately drift

together, owing to their mutual attractions, and so neutralise one another, the strain forms constituting the charges moving across the ether without sensibly interfering with its motion. The consequences of long-continued accumulation will therefore be obviated by the action of a process which is at each instant so insignificant as to be far below the reach of observation.

These considerations may be regarded as completing the formal development of the theory, and the following chapters will be mainly occupied in its application to various classes of physical phenomena, not necessarily confined to those observable on our own planet.

We have arrived at the conclusion that the basis of the visible material universe must be regarded as consisting of a uniform continuous medium associated in every smallest portion with energy of spin. This energy of spin confers upon it the property of elastically resisting distortion having the nature of twist, but it offers no resistance to twistless flow. The medium contains two kinds of strain form differing only *chirally* from each other; that is to say, their relation to each other is similar to that of an object and its image in a plane mirror, or to that of a right-handed and a left-handed screw. These strain forms are capable of flitting freely through the ethereal medium in which they are formed without sensibly affecting its motions. It is by the aggregation into groups of these strain forms, usually in orbital motion relatively to one another, that all material bodies are formed. These groups will be capable or incapable of unresisted motion through the ether according as they do not or do give rise to ethereal twist as the result of their motion. This ethereal twist is known as electric displacement, and is necessarily associated with the velocity of flow of which it is half the curl, and this velocity of flow is known as the magnetic induction. The only force concerned in the transmission of action across the free ether is electric force, determined by dividing the twist, known as electric displacement, by a coefficient representing the elastic resistance of the medium to twist. The sole origin of this twist, and consequently of electric force, is the motion of groups of strain forms through the ether. When ether twist is set up in a region containing strain forms, its distribution is modified by their presence, resulting in what are known as pondero-motive, or mechanical, forces, tending to set the groups of strain forms in motion. The grouping of the strain forms existing in the material substances known as conductors allows some of the forms to move freely, except for mere viscous resistance, through the material,

giving rise to electric currents. The grouping of the strain forms existing in the material substances known as dielectrics, on the other hand, elastically resists their displacement from their original positions, so that the distance traversed is limited, and depends on the amount of the ether twist, giving rise to electric polarisation. In a system consisting only of material dielectrics and free ether, a steady ether twist is counterbalanced by the elastic resistance called into action, so that the application of such a steady twist will, after a few oscillations, give rise to a state of static equilibrium. The application of a continuously alternating twist will give rise, in such a system, to a series of electromagnetic oscillations or waves; that is to say, to radiation. When, however, the system contains conductors, the strain arising at the surface and within the substance of the conductors, from the application of a steady twist, will be continually relieved by the motion of translation of the free strain forms, and the electric currents so arising would continue for ever were it not for the viscous resistance to the motion of these strain forms, causing a dissipation, or transformation of energy into heat. A continuous supply of energy is therefore necessary for the continued maintenance of the current.

The electron theory, thus briefly summarised, enables us to form a far more complete mental representation than has hitherto been possible of the circumstances of an electric field, that is to say, from the present point of view, of very nearly the whole complex of physical phenomena. It does not, however, enable us to form a conception of any direct mechanism by which the pondero-motive forces are transmitted across the ether from one conductor to another. We are accustomed in material mechanisms, which are the only ones of which our experience can enable us to form definite mental pictures, to consider the forces acting between material bodies as tractions over definite surfaces bounding the matter. From our present fundamental point of view, the forces acting on matter are forces acting on strain forms spreading outwards from their nuclei throughout the whole medium, so that the matter cannot be considered as possessing any definite boundaries. We can form, comparatively speaking, a fairly satisfactory mental picture of the transmission of electric force by means of ether twist through ether free from strain forms, but even here we must remember that the forces so transmitted are electric forces and not mechanical tractions, so that our conception of a sort of wheelwork is only a crude analogy, and cannot be considered in any fuudamental sense as representing the actual transmission system. When matter, that is to say, groups of strain forms, is present, the distribution of force depends partly on the relative configuration of the strain forms in the ether. These strain forms have, in

the strict sense of the term, potential energy; that is to say, energy of position due to their configuration, and when this is altered, the work done by the forces acting between them may be employed, either in accumulating other potential energy elsewhere, or in increasing the kinetic energy of the matter, this kinetic energy itself consisting of energy in the ether arising from the movement of the strain forms across it.

The only conceivable method of dealing with such a problem is to determine the mathematical function representing the total energy of strain, and to derive from this function the mutual forces, or stress relations, by the application of the principle of the conservation of energy, that is to say, the negation of perpetual motion. Explanation of the interactions cannot possibly have any other meaning than the carrying out of this process, and any mental pictures based on material analogies, helpful as they may be as aids in correlating the various phenomena, can never be regarded as more than crude analogies, or, in any valid sense, as constituting explanations of them.

Sir Joseph Larmor points out that similar considerations apply to the vortex-atom illustration of matter :—[1]

"If we consider rigid cores round which the fluid circulates, they are moved about by the fluid pressure; but if we consider vortex rings, say with vacuous cores, these are mere forms of motion that move across the fluid, and if we take them to represent atoms, the interactions between aggregations of atoms cannot be traced by means of fluid pressures, but can only be derived from the analytical character of the function which expresses the energy."

---

[1] *Phil. Trans.*, vol. cxc., A, 1897, p. 214.

CHAPTER X.

MAGNETISM AND THE DISSIPATION OF ENERGY.

In the original form of the Ampère-Weber theory of magnetism, that is to say, of paramagnetic action, it was assumed that the molecules of a magnetic substance in the unmagnetised condition are oriented indifferently in all directions, and that the directing force, acting as a constraint on the turning of the molecules under the action of a magnetic field, is due to a magnetic force of some constant value acting in the primitive direction of the magnetic axis of the molecule. Various modifications of this directing force were suggested by Maxwell, Wiedemann, and others, but Ewing showed, in 1890, that no such artificial hypothesis was required, as all the phenomena could be far more adequately accounted for by the mutual magnetic actions between the molecules themselves. Before this, Hughes had shown good grounds for believing that the molecules are not indifferently oriented in all directions prior to magnetisation, but that they are so grouped into aggregates that the external effect of each aggregate is practically null.

Weber's theory of diamagnetism ascribed it to currents in the molecules caused by external electromagnetic action upon molecules in which there were no permanent molecular currents such as were assumed in order to account for paramagnetic phenomena. In the case of strongly paramagnetic substances, the effect of these induced currents was supposed to be so completely masked by that of the permanent currents as to be insensible to observation. On the electron theory, there must be permanent currents, arising from the orbital motions of the electrons, in all material atoms. It is quite possible that there may be considerable differences in the strengths of these permanent currents in the atoms of different chemical elements, but it is hardly conceivable that the very marked paramagnetic effects observed in iron, nickel, and cobalt should be entirely or even largely due to such differences. The observed phenomena point rather to peculiarities of molecular constitution as the cause of *ferromagnetism*, as the

exceptional development of paramagnetic properties observed in these substances is called.

We are therefore led to a modification of Weber's theory, in which paramagnetism is accounted for by the *orientation* of the permanent molecular currents, that is to say, of the orbits of the electrons of which the atoms are composed, under the influence of an imposed magnetic field, while diamagnetism is accounted for by the *alterations* in the orbits arising from the disturbing action of the imposed field. The exhibition, to some extent, of both paramagnetic and diamagnetic effects is therefore to be looked for in all substances, and marked magnetic peculiarities, more particularly those exhibited by ferromagnetic bodies, must be supposed to be mainly due to peculiarities of molecular constitution. The diamagnetic effects resulting from disturbances of the orbits, though individually very minute, will be cumulative; while the orientation representing the paramagnetic effects will be a differential one, being of opposite character, according to whether the direction of orbital motion is clockwise or counter-clockwise.

The phemonena of induced magnetisation in the strongly paramagnetic substances, iron, nickel, and cobalt, are of a very much more complex character than was suspected at the time that Maxwell's treatise was written. In the magnetisation of any of these metals the magnetising process may be broadly divided into three stages, indicated by A, B, and C respectively, in fig. 18, which is a typical example of the susceptibility curve $I = \kappa H$,[1] or the curve of magnetisation, as it is usually called. During the first stage, A, corresponding to very small values of H, the curve has a very easy gradient, indicating a comparatively small susceptibility. During the second stage, B, the curve is very nearly a straight line, and the gradient is steep, indicating a high and very slowly increasing susceptibility. As H is further increased, the gradient eases off again, with diminishing susceptibility, and ultimately becomes practically indistinguishable from a straight line. This line very clearly indicates an approach to a limiting value of I, which is known as the *saturation* value. It is impossible to say that the approach to this limit is not asymptotic, but in wrought iron it is practically attained before H has risen to 2000 C.G.S., after which the magnetic force may be increased tenfold without sensibly increasing the magnetisation.

[1] The magnetisation I may be measured by observing the deflection of a magnetometer needle. In the case of sudden changes in I it is preferable to employ the *ballistic* method, in which the change in I is determined by measuring the transient current which is induced, during the change, in a coil surrounding the magnetised piece. (See J. H. Ewing, *Magnetic Induction in Iron and other Metals*, chapters ii. and iii.

Experiments made by Lord Rayleigh in 1887 show that, for very small values of H, κ is constant for iron, and there is no *retentiveness* and consequently no *residual magnetism*; that is to say, when the magnetic force is removed, the induced magnetisation disappears. He also examined the effect of alternately applying and removing small magnetic forces to and from an iron bar kept magnetised by a constant magnetic force, and found that, when this constant value of H is small, the value of κ is sensibly the same as its initial value, but as this constant value is increased in amount, the susceptibility with respect to small changes is gradually reduced. Under the action of these small forces, the magnetisation, especially in the case of large bars of soft iron, may

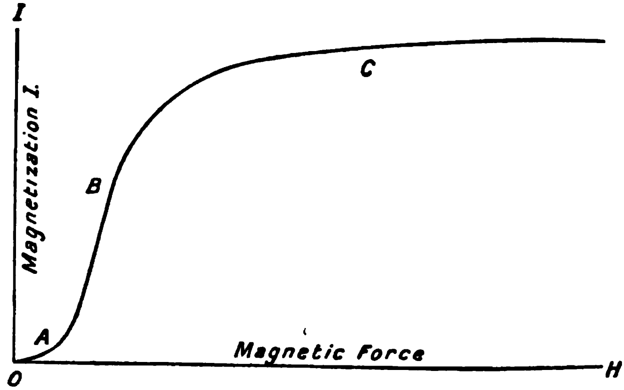


FIG. 18.

not attain or lose its full value for several minutes after the application or cessation of the magnetising force, so that there is a time-lag of the magnetisation behind the magnetic force. In hard iron and steel this time-lag is so small as to be scarcely perceptible. It may be accounted for by a frictional, or viscous, resistance offered to the turning round of the molecules, being analogous to the similar time-lag observed in dielectrics.

When the magnetic force attains higher values, the magnetisation no longer disappears entirely when the magnetic force is removed, but tends to persist, causing the magnetisation to lag permanently behind the magnetic force—a phenomenon to which Ewing has given the name of *magnetic hysteresis*. The first and most obvious result of this is that the demagnetisation curve differs considerably from the magnetisation curve. Moreover, if the metal is again subjected to the action of an increasing magnetic

force, the new magnetisation curve will differ slightly from the original one, and the final value of the magnetisation attained when H has regained its initial maximum value is in some cases found to be slightly higher. If this process is repeated a sufficient number of times, the curves of magnetisation and demagnetisation will each attain constant forms differing from each other. That is to say, the process of magnetisation and demagnetisation ultimately becomes *cyclic*. Hysteresis may be diminished by mechanical vibration, or by any other disturbances which make the metal readier to respond to influences tending to alter its magnetism ; such, for example, as increase of temperature, up to a certain extent, after which other considerations come into play, as we shall presently see.

These phenomena are all in accordance with the suggested modification of the Ampère-Weber theory, but, as Ewing has shown, they will be accounted for on any theory of magnetic molecules originally grouped into molecular aggregates which may be more or less broken up by increasing the magnetic force, especially when the effect of this increase is supplemented by mechanical or other disturbance of a character tending to promote rupture of the ties in these molecular groups. In the first stage A we may suppose that none of these groups are broken up, so that on the removal of the magnetic force, the quasi-elastic stress, arising from the deflection of the molecules, restores them to their initial positions. Moreover, if at any stage of the magnetising process we cease from increasing H and begin to decrease it, or cease from decreasing H and begin to increase it, the rate at which I will change relatively to H will evidently be very small at first, depending, as it does, only upon this quasi-elastic movement of the deflected molecules. It will only be after the reversal has been carried to some little extent that it will begin to affect the stability of the molecular groups. The second stage B will begin when H reaches a sufficient value to begin breaking down the magnetic ties between the various molecular groups. The sharp gradient of the curve shows that the breaking up of the groups occurs slowly at first, and proceeds with increasing rapidity as H continually increases, corresponding to a rapidly increasing susceptibility. A bend, or knee, is then reached in the curve, after which the susceptibility begins to diminish with gradually increasing rapidity, forming the third stage C. Even during this stage the breaking up of the molecular groups continues, as is shown by the fact that the curves of increasing and decreasing magnetisation do not become indistinguishable until I ceases to show sensible increase with further increase of H, and the curve becomes sensibly parallel to the line OH.

Ewing has obtained experimental confirmation of the validity of this explanation of the observed results by studying the behaviour of groups of small magnets in increasing and decreasing magnetic fields. The magnets were made of short strips of hard steel, pivoted, like compass needles, on fixed centres, the strips being slightly bent at the centre so as to bring the centre of gravity below the point of the pivot, and a recess for this point being stamped by a centre punch in the hollow of the bend. The pivots were formed of needles, inserted, with the points upwards, in small blocks of lead or of wood. These magnets, being of hard steel and strongly magnetised, have their magnetic moments, and therefore their magnetic intensities, practically unaffected by the magnetic fields applied to turn them into line. A magnetisation curve obtained by Professor Ewing with four such magnets is shown in

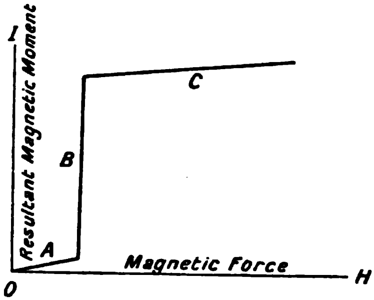


FIG. 18A.

fig. 18A. When the number of magnets was gradually increased up to a couple of dozen, the suddenness of change from one stage of the curve to the next diminished, showing a steady approach to a continuous curve of the character shown in fig. 18. Ewing states that he has obtained in this manner very close representations of the phenomena attending the reversal of magnetism, the dissipation of energy in hysteresis, the conditions that promote residual magnetism, the comparative effects of slow and sudden changes in magnetic force, the effects of mechanical strain, the influence of vibration, and the existence of time-lag.

Further confirmation of the validity of Ewing's explanation is afforded by the fact that the hysteresis vanishes when iron is rotated in a very strong magnetic field, as was shown by Baily in 1896. James Swinburne had previously pointed out that this result would necessarily follow from Ewing's theory, as when the field was sufficiently strong, each elementary magnet would simply turn with the field, preserving its direction along a line of force,

and would therefore form no ties with contiguous molecules, and have no tendency to be set into oscillation.

It was known to Gilbert that iron and steel sensibly lose their magnetic properties when heated to bright redness, but recover their susceptibility on cooling. The same thing occurs at a lower temperature with nickel, and at a higher temperature with cobalt. Dr John Hopkinson, who investigated this subject in 1888 and 1889, calls the temperature at which a metal sensibly loses its magnetic properties the *critical temperature*. He finds that, in samples of ordinary iron and steel, this varies from 690° C to 870° C. Many other phenomena, which had previously been noted by various observers, are all found to occur at this same critical temperature, and show that some fundamental change in molecular constitution must be taking place. It was observed, for example, by Gore, in 1869, that an iron wire, when cooling from a bright red heat, undergoes a sudden elongation at a dull red heat, and then continues contracting as before. Tait observed, in 1873, a sudden change in the thermo-electric quality of iron at this temperature; and both Hopkinson and W. Kohlrausch have shown that the effect of temperature on the electrical resistance of iron undergoes a sudden change in amount at the critical temperature. Finally, Barrett observed, in 1874, that iron which is cooling slowly through a temperature—which he suggested might be found to be the one at which it regains its magnetic properties—has its cooling suddenly arrested, and sometimes even has its temperature momentarily increased, giving evidence of an internal liberation of heat which can only be accounted for by a change in the molecular constitution. In hard steel there is a perfectly visible re-glow at this point. The phenomenon is known as *recalescence*. Hopkinson showed that the critical temperature was also that of recalescence. He then measured the heat evolved, and showed that a corresponding absorption of heat takes place during the passage from the magnetic to the non-magnetic state. He also showed that it did not occur at all in non-magnetisable manganese steel, a further indication of its intimate connection with the magnetic conditions. Rowland showed in 1874 that nickel at 230° C. had a higher susceptibility to weak values of H than at 5°, but for higher values of H this effect was reversed, the susceptibility being less at the higher temperature. He showed that cobalt behaved in a similar manner with regard to weak magnetic forces, and it has since been shown that a similar reversal in the effect of temperature takes place when the value of H is sufficiently increased, this value being considerably higher than in the case of nickel. Baur, in 1880, showed that iron behaved in a similar

manner. Hopkinson's more complete investigations gave some very interesting results. For $H = 0{\cdot}3$ C.G.S. the susceptibility increases, slowly at first, until a temperature of 600° C. is passed. It then rapidly increases more than tenfold, showing that the effect of the heating is to bring on the second stage of the process of magnetisation. The increase continues up to 775° C., and then falls suddenly until the critical temperature of 785° C. is reached, when the susceptibility sensibly disappears. For $H = 4$ C.G.S. there is no such sudden rise in the susceptibility, the second stage in the magnetisation having then already been entered upon; the loss of susceptibility at high temperature is more gradual, and this diminution in the rate of decrease of the susceptibility becomes still more marked for higher values of H.

From the time of Gilbert's observation of the sensible disappearance of the magnetic properties of iron at a bright red heat, until so late as the year 1895, it was supposed that its demagnetisation, and the subsequently discovered demagnetisation of nickel and cobalt at their own critical temperatures, was complete. In that year, however, P. Curie, as the result of a very extensive investigation of the magnetic properties of matter, carried out by very sensitive methods, which was published in the *Annales de Chimie*, arrived at the conclusion that in all feebly paramagnetic substances, including gases, the magnetic susceptibility is inversely proportional to the *absolute temperature*,[1] to a degree of approximation approaching perfection at high temperatures, and that, at sufficiently high temperatures, even such strongly magnetic substances as iron, nickel, and cobalt, which are known as *ferromagnetic*, ultimately follow the same law, which is known as *Curie's Law*, but that in diamagnetic substances the susceptibility is usually almost independent of temperature, and also of changes in the chemical combination of the atoms. The inference is drawn by Curie that diamagnetism is a phenomenon depending mainly on the internal constitution of the molecules, and having only slight relation to their bodily motions, on which the temperature depends, while paramagnetism is dependent entirely on the orientation of the molecules in space, and not upon their internal constitution. According to this view, paramagnetism would necessarily be affected by changes in the average bodily motions of the molecules, so that a temperature effect is to be expected, but no such effect is to be expected upon diamagnetic action.

These results are completely accounted for by the modified Ampère-Weber theory which is involved in our present view of the constitution of matter, according to which both paramagnetic and diamagnetic action is to be looked for in every kind of matter,

[1] See Appendix I.

the one or the other preponderating, according to the manner in which the molecules of each substance are built up from their constituent atoms. From this point of view the slight traces of temperature effects observed by Curie in diamagnetic substances are to be attributed to the paramagnetic action which can never be entirely absent. Even in feebly paramagnetic substances, the paramagnetic action is so preponderant as to almost completely mask any diamagnetic action; and according to Curie, his experimental results in case of paramagnetic substances may be equally well represented by a formula taking account of the paramagnetic action only, or by one in which the diamagnetic effects are included as well.

Curie calls attention to the strikingly close analogy existing between the simple law of magnetisation of ferromagnetic substances above their critical temperatures and the sudden change in its character when the temperature falls below this point; and the correspondingly simple law of expansion of a substance in the gaseous state at a high temperature, and its sudden alteration in character when the temperature is lowered beyond a critical point at which the mutual attractions of the molecules come into play, and result in liquefaction.

Sir Joseph Larmor observes that the relation between paramagnetisation and temperature, at temperatures above the critical one, is so simple and universal that it must be the expression of a theoretical principle; and he shows that when the conditions are such that the mutual magnetic actions of the molecules are insensible, Curie's law may be deduced directly from Carnot's principle.[1] Conversely, if Curie's law be assumed, it will follow that the magnetisation of a paramagnetic substance must consist simply of orientation of its molecules without sensible change in their internal energies.

An ordinary paramagnetic substance must therefore be regarded as one in which the mutual magnetic control of the molecules is insensible in comparison with the control due to other molecular actions, and which are of such a character that the magnetic energy expended in overcoming them is transformed into heat, and not into internal energy of any elastic type. A ferromagnetic substance, on the other hand, must be regarded as one in which the controlling force is almost entirely derived from the mutual magnetic actions of neighbouring molecules.

We are now in a position to account for the relations observed by Hopkinson and others between temperature and the magnetic qualities of ferromagnetic substances. The effect of an increase of temperature in a weak magnetic field, in accelerating the transition

[1] See Appendix J.

from the first to the second stage of magnetisation, is accounted for by the breaking of the ties between the molecular groups, due to the combined effects of expansion and increased molecular agitation. During the later portions of the first stage of magnetisation the metal is in a critical state, in which a small increase in the magnetic force will largely increase the susceptibility by breaking up groups which were stable in the weaker magnetic field. It is quite comprehensible that, under such circumstances, the decrease in stability of these molecular groups, owing to the rise in temperature, should result in very greatly accelerating the breaking-up process, and so greatly increasing the susceptibility. A stronger field is in a less critical state, many of the groups being already broken up, so that the effect of the temperature in diminishing the susceptibility, owing to the increased oscillations virtually reducing the average magnetic moment of each molecule, more nearly counterbalances its effect in breaking down the grouping, until finally the former effect becomes predominant, and the curve bends downwards.

It was pointed out on page 131 that magnetic permeability may be considered as the analogue of specific electric conductivity, and Faraday actually called it the *conductivity for magnetic lines of force*, an idea which Hopkinson developed into the conception of a magnetic circuit, which has been largely employed in the solution of practical problems in connection with electrical machinery. We saw in Chapter IX. that the electromotive force of an electric circuit is measured by the amount of work which would be done in carrying unit quantity of electricity completely round the circuit, or, what comes to the same thing, it is the line integral, round the circuit, of the electric force E. Similarly, what Bosanquet has called the *magnetomotive force* round a magnetic circuit is measured by the amount of work which would be done in carrying a unit magnetic pole completely round the circuit, or, it is the line integral, round the circuit, of the magnetic force H. Suppose, for simplicity, that the circuit is one for which the susceptibility κ and the cross section s are the same throughout. Then the flux across any section of the circuit is Bs, and $B=\mu H$. Moreover, in this case the magnetomotive force will be $\bar{H}l$ where l is the length of the circuit and $\bar{H}$ is the average value of H round the circuit, and therefore

$$\frac{\text{magnetic flux}}{\mu s}=\frac{B}{\mu}=\bar{H};$$

therefore,

$$\frac{\text{magnetic flux}\times l}{\mu s}=\bar{H}l,$$

or,

$$\text{magnetic flux} = \frac{\bar{\text{H}}l}{\dfrac{l}{\mu s}} = \frac{\text{magnetomotive force}}{\text{magnetic resistance of circuit}}, \text{ say,}$$

which is in the same form as the equation of electric conduction expressing Ohm's law, viz.,

$$\text{electric current} = \frac{\text{electromotive force}}{\text{electric resistance of circuit}}.$$

If R is the resistance of the electric circuit, r its specific resistance, s its cross section, both supposed uniform throughout, and l its length, then $\frac{lr}{s} = \text{R}$, and $\frac{1}{r}$, the specific conductivity, being taken to correspond to μ, we see that we must take the magnetic resistance, as above, as being equal to $\frac{l}{\mu s}$.

This analogy must not be pushed too far, for μ is not, like $\frac{1}{r}$, a constant quantity for a given material, but varies with the value of B, and depends also on the previous magnetic history of the material.

This point of view is, however, of interest in connection with our present theory, which regards the tubes of magnetic induction as tubes of actual twistless ether flow, and it may be illustrated by figs. 19 and 19*a*, which have been copied from a paper of Lord Kelvin's.[1] Fig. 19 shows the disturbance produced in an originally uniform magnetic field by the introduction of a soft iron sphere. Before the introduction of the sphere, the lines of magnetic induction were straight, parallel, and equally spaced. They might then represent the lines of flow of a liquid, such as water, for example, all portions of which were flowing with the same speed and in the same direction. Let us suppose, for example, that they represent the lines of flow in a portion of a straight, slow-moving stream, partially, but uniformly, obstructed by a mass of growing weed, the portion considered being so far removed from the banks and the bed of the stream that its motion is uniform throughout. The distortion shown in fig. 19 could then be imitated by maintaining a central spherical space clear from weed, so as to allow a freer flow through this portion than elsewhere. Fig. 19A shows the effect of introducing a sphere of strongly diamagnetic material into a magnetic field which was originally uniform. In this case we may suppose the original straight, parallel, and equally spaced lines to represent the lines of flow in

[1] *Papers on Electrostatics and Magnetism*, pp. 489 and 491.

an unobstructed portion of the stream. The distortion might then be imitated by introducing into the centre of the flow a partial obstruction in the shape of a spherical mass of uniformly distributed weed. The dotted lines, which are not shown in the original paper, illustrate the fact that in each case the magnetic induction within the sphere is uniform.

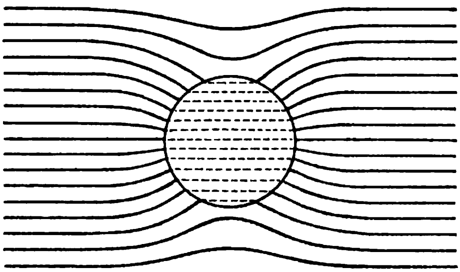


FIG. 19.

A very important consequence of the phenomenon of magnetic hysteresis is that all changes in the magnetisation of ferromagnetic substances, with the exception of such very small changes as disappear entirely on the removal of the magnetising field, so that no alteration in the molecular grouping occurs, involve a dissipation of energy by its transformation into the *unorganised* form of

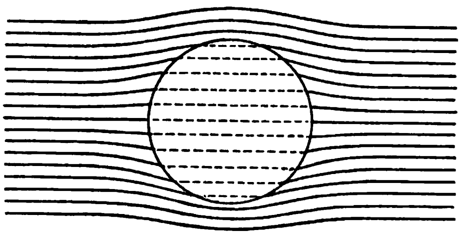


FIG. 19A.

heat energy. When the magnetism is carried through a complete cycle of changes, the hysteresis curve expressing the relation between H and I will be a closed one, and Warburg showed in 1881 that its area[1] measures the amount of energy dissipated during the cycle on account of hysteresis. Ewing discovered

[1] Viz., the value of the integral $\int H dI$ taken round the curve.

the same fact independently in 1882, and a simple proof, due substantially to Hopkinson, will be found on page 99 of his work on *Magnetic Induction in Iron and Other Metals.*

This dissipation of energy by hysteresis constitutes a considerable proportion of the energy losses in many kinds of alternating current machinery, especially in transformers. A very interesting phenomenon, first observed in transformers and described by G. W. Partridge in 1894, is the gradual increase which takes place in the hysteresis of the iron cores, accompanied by a diminution in the permeability of the iron in weak magnetic fields. W. M. Mordey showed in 1895 that these effects were due to the prolonged heating of the iron. Roget found, in 1898, that even a temperature as low as 50° C., if continued for several weeks, produces an appreciable effect. In one of the samples of transformer iron experimented on, a temperature of 160° C. increased the hysteresis to double its original amount in the course of a few hours, and to nearly three times the original amount in the course of a few days. When the heating is further prolonged at such high temperatures, the hysteresis attains a maximum, and then begins to diminish again, but without ever falling to the initial amount previous to heating. The effect is evidently due to molecular changes in the metal, and as very similar changes occur in iron and steel subjected to continuous mechanical vibration, it probably arises from the increased oscillations of the molecules at high temperatures. Moreover, Ewing showed, in 1886, that structural changes involving hysteresis occur when iron is subject to cyclic variation of stress, even in the absence of magnetic force. He found, for example, that when an iron wire under tension is loaded and unloaded, by putting on and taking off weights, there is a distinct difference in the physical state of the metal, under one and the same intermediate amount of weight, during loading and during unloading, the difference showing itself both in magnetic susceptibility, and, as was shown by the present writer (*Phil. Mag.*, 1878), also in thermo-electric quality.

It was pointed out by Helmholtz in 1881 that in any polarised medium there must be a material tension along the lines of polarisation and a pressure at right angles to them. Each of these is, as Sir Joseph Larmor showed in 1892, proportional to the square of the susceptibility, and they need not be equal. For media of high susceptibility they are, therefore, far more intense than Maxwell's hypothetical stress, which is directly proportional to the susceptibility. Larmor illustrates their origin by the case of a chain of iron nails hanging end to end from a pole of a magnet; the nails hold together longitudinally, but repel each other

transversely. In a longitudinally magnetised rod, this straining together of opposite polarities in neighbouring molecular groups will give rise to an internal stress in the bar proportional to I^2, where I is the intensity of the magnetic polarisation, which will usually tend to shorten it, but may conceivably lengthen it. The orientation of molecular groups due to magnetisation will directly alter the length to an extent proportional to I, so that the whole increase of length will be represented by the formula $AI + BI^2$, where A and B are constants for a given material, and may each be either positive or negative. When A and B have opposite signs there will be a value of I at which the total effect will reverse its sign by passing through a null value. This is found experimentally to be the case for iron, in which A is positive and B negative, and for cobalt, in which A is negative and B positive. For nickel, A and B are both negative.

Larmor points out that there must be similar strain effects, though they would be far more difficult to detect, connected with the polarisation of a dielectric in an electric field; and he considers that the main part of the phenomena of *electrostriction* and *magnetostriction* is of this character, and that only a comparatively small portion is due to strain of the material due to the direct effect of electric and magnetic attractions between finite portions of the material.

The system formed by a bundle of iron nails suspended from a pole of a magnet will afford a very good representation of a polarised medium, either magnetic or dielectric, if we assume with Larmor that the nails are connected with one another by means of springs, or are embedded in an elastic matrix. When there are no external forces acting on such a model, it will adjust itself into a condition of internal equilibrium in which the attractions between the magnetic nails are locally balanced by the elastic forces due to the springs or to the matrix. The various local systems of molecular forces, or molecular *forcives*, represented by attractions between magnets, and by elastic forces, will exactly balance one another in every portion of the medium. The introduction of an additional magnetic field will cause changes in the magnetic polarities of the nails, so that the different parts of the medium will change their shapes and volumes until equilibrium is again attained. There will therefore be an intrinsic deformation of the medium, and possibly also intrinsic changes in its other physical properties, associated with the polarisation. The polarisation in an element of volume being measured by the product of the number of polarised molecules contained in the element of volume, into the average value of the magnetic or electric moment

of the molecules, these changes will, in simple cases, be proportional to the square of this product.

Now suppose that an external force, such as gravity, or the magnetic field arising from the medium as a whole, begins to act; this will be a regular mechanical force acting on the medium in bulk, and will therefore, in the aggregate, be proportional to the volume on which it acts. This will give rise to further deformation, but one proportional directly to the exciting force, instead of to its square. The local internal molecular forcive will then be no longer exactly balanced. The unbalanced part will, however, everywhere form a stress system similar in character to that existing in a strained elastic solid, because, when the element of volume is small enough, the resulting tractions over its surface must balance one another independently of the external forcive, which will in that case be negligibly small. If the dimensions of the smallest element of volume which the circumstances of any special case make it necessary to consider, are large in comparison with the range of the molecular forces, the elastic stress due to external forces will then be capable of expression in terms of tractions over the surfaces of these elements of volume.

This separation of the forcive per unit volume of a polarised medium into distinct molar and molecular portions finds its earliest exemplification in Poisson's ideal volume and surface densities, sometimes employed by Maxwell, and the effect of which is equivalent at external points to that of the actual polarisation. Larmor observes that

> "This method consists essentially in computing the forcive by combining opposed poles of neighbouring elements, instead of taking the single polarised element as the unit; it shows that these adjacent poles nearly compensate each other, except as regards a simple volume density whose attraction has no molecular part, and a surface density, partly at the outer surface and partly at the surface of the cavity which contains the point under consideration. The effect of the latter surface density, depending as it does wholly on the immediate surroundings, is the molecular or cohesive part of the average forcive."[1]

The justification of this theory, according to which the molecular part of the forcive due to polarisation is balanced locally by an intrinsic cohesive stress in the material, which is independent of the material elastic constants and strains at

[1] *Phil. Trans.*, vol. cxc., A, 1897, p. 258.

the place, is further enforced by Larmor by considering the ideally simple case of a gas. A system of bodily forces acting from a distance on a mass of gas can always be balanced by a simple increase of pressure, provided the forces are derivable from a potential, and if they were not so derivable the medium could not be in equilibrium. In this case we know, from the kinetic theory of gases, that the internal forcive acting on an element of volume is mainly due to the impacts of the molecules surrounding the element, and partly also, in the case of dense gases, to cohesive molecular actions. The extraneous forcive by which this internal forcive is balanced is one acting from a distance by means of the ether. The balance is therefore one between a static bodily forcive and a steady kinetic molecular one, and if the distant bodily forcive is altered, equilibrium can only be maintained by corresponding adjustments in the impacts and the local molecular attractions. When co-ordinated electric or magnetic actions are added to the bodily forcive by polarisation of the gaseous molecules, a further adjustment of molecular configuration must take place. Now, the forcive due to polarisation is not isotropic like a fluid pressure, but depends on the direction of polarisation. Its action will therefore alter the originally indifferent distribution of the velocities of the gaseous molecules into one of a slightly axial character, and when equilibrium has again been attained, the pressure due to the impacts on any element of surface will vary with the inclination of the element of surface to the direction of polarisation. This additional intrinsic local stress must exactly balance the action of the local electrical or magnetic actions when the gas has again attained a condition of equilibrium, and there will therefore remain nothing of the electric or magnetic mechanical forcive to be transmitted across the gas. The argument in a suitably modified form may be extended to any isotropic medium. In the case of an ælotropic solid, the molecular stress balancing the electric or magnetic forcive on each element will be of a more complex character, as it will depend on the axes of ælotropy as well as on the axis of polarisation.

We arrive, therefore, generally at the very important conclusion that, under conditions of equilibrium, that part of the forcive on an element of a material body which arises from electrical or magnetic excitation of neighbouring mole-

cules, and is expressed in terms of them alone, is not transmitted by material stress, but forms a balance locally with the internal molecular forcives of other types.

It follows, therefore, that a dielectric in an electric field, and any material body in a magnetic field, will experience a mechanical strain due to the field, and that it will also undergo intrinsic changes of volume, shape, and other physical properties, proportional to the square of the local polarisation. In the case of solid bodies, these changes of dimensions may give rise to secondary strains; but these cannot arise in fluid media, in which no strain other than simple compression is possible. This principle, restricted, however, to non-polar forces, was first enunciated by Thomas Young in 1805 in a paper in the *Philosophical Transactions of the Royal Society*, "On the Cohesion of Fluids," and applied to the consideration of capillary phenomena.

Larmor observes:—[1]

"The scope of these molecular considerations is wider than the special problem of polarisation by which they are here precisely illustrated. To an intelligence that could follow the play of interaction between the individual molecules of matter, mechanical forces, in the ordinary sense, would not exist. The actual interactions between the molecules are, however, necessarily presented to us divided into various statistical groups, which are the subjects of perception by different senses; and it is the business of physical theory to follow out the relations of these different groupings to each other, and to trace them all back into the ultimate unity. The total energy of the molecules of a material body, corresponding to any kind of excitation or polarisation, is thus for us made up of various parts. There is a part involving the interaction, with any molecule under consideration, of other molecules at finite distances, which integrates [that is to say, sums up] into an energy function of applied mechanical forces of the system, such for example as gravitational or magnetic forces. Of the remainder of the energy, which arises from the mutual actions of neighbouring molecules, a regular, or organised, part can be separated out which represents the energy of elastic stress and is a function of the deformation of the element of volume treated as a whole. This stress, arising from the immediate surroundings, in part compensates, for the element of mass under consideration, the applied mechanical forces aforesaid. The remaining, usually wholly irregular, parts of the local intermolecular forces and motions compensate themselves mutually on the spot—or at any rate can be considered as thus compensated by other such forces, of different origins, that are not at present under consideration."

Larmor points out that this order of ideas is necessarily

[1] *Loc. cit.*, p. 260.

involved in d'Alembert's principle, which forms the basis of the dynamics of material bodies.[1]

"That part of the aggregative forcive on the molecules in the element of volume which is spent in accelerating the motion of that element *as a whole*, is written off; and the regular part of the remainder must mechanically equilibrate. But the wholly irregular parts of the molecular motions and forces are left to take care of themselves; which they are known to do, for the simple reason that the constitution of the material body is observed to remain permanent."

The temperature depends, in fact, on this irregular residuum of forces, and so do the density and the other physical properties of the medium, which are thus affected when, owing to polarisation or other excitation, the local part of the molecular forces and motions is altered.

"If we agree to maintain the original precise meaning of the term mechanical (as above employed), viz. that a mechanical force is one which we can actually control for doing work for our purposes on matter in bulk, in contrast with a molecular force which we can reason about but not directly employ, we may call the regular part the *mechanical* energy and the remaining wholly irregular part the *non-mechanical*; we may also use (as above) the terms *organised* energy and *unorganised* energy with the same meaning, the reference being now to the material medium as a continuous organic whole, transmitting applied forces by stress, not as a numerical aggregate of separate molecules."

The distinction thus made must not be considered as exactly equivalent to Lord Kelvin's two classes of *available* and *dissipated* energy, for unorganised energy may frequently become partly transformed into organised energy by means of physical transformations involving sifting processes of molecular fineness which are necessarily non-mechanical. For example, as Lord Rayleigh showed in 1875, the unorganised energy of two masses consisting of different gases, at equal temperatures and pressures, may be in part converted into organised energy, and therefore into mechanical work, by allowing transpiration to take place between them through a porous plug or partition, the diameters of whose pores approach molecular dimensions. The gases will then to some extent sort themselves, owing to the average value of the squares of the velocities being different for the two gases.[2] The pressure on one side of the partition will therefore become greater than on the other, so that if the partition were movable like a piston, work might be done by pushing

[1] *Loc. cit.*, p. 261.
[2] See Poynting and Thomson's *Heat*, chapter ix.

it in the direction of the mass of gas in which the pressure had diminished, and the operation would be accompanied by a fall in temperature of the whole quantity of gas, the mechanical work being derived from some of the heat energy of the gas. Again, the stores of energy of chemical combination existing in an unorganised form in electrolytes can be largely transformed into organised energy, and therefore utilised, by the sifting, or separating, action of electric force on their ionic constituents.

The unavailable energy of Lord Kelvin must be regarded as the *residuum* which we cannot render mechanical by any sifting process in bulk, but only by means of actual constraint applied to the individual molecules. Forms of energy which are at present considered to find their place entirely in this residuum may prove to be partly reducible to a mechanical form as our knowledge of physical transformations increases. On the other hand, however, the recognition of temperature effects in processes previously considered reversible shows that some forms of energy formerly regarded as wholly available are really only partly available.

The principle of the dissipation of energy was recognised and pointed out by Lord Kelvin in 1852, as a consequence of the fact that, while the whole of the energy of a mechanical system may be converted into heat, the inverse operation can only be partly effected, the utilisation of the whole of the heat in performing mechanical work being an impossibility. According to the principle of the conservation of energy, the total energy of any isolated system must remain unchanged. It will, however, of its own accord—that is to say, without the application of any external force—tend to pass through transformations which will diminish its availability; for, since work is done only when a transformation of energy occurs, every change undergone by the system of its own accord, and in which therefore any work done is unutilised, diminishes its capacity for doing useful work.

Although all the phenomena of which we have any experience are subject to this principle, it is not of the same fundamental character as that of the conservation of energy, which can hardly be conceived as being otherwise. Clerk Maxwell has shown that the influence of the dissipation principle over our actions is merely a consequence of the rudeness of the machinery at our disposal for controlling the behaviour of those portions of matter to whose relative motions or positions the energy we have to deal with is due. He supposes a vessel containing a gas at a uniform

temperature throughout to be divided into two compartments A and B by a partition containing a number of doors capable of being opened and closed without any expenditure of energy. An intelligent creature, which we may call a "sorting dæmon," and which is supposed to have an infinitely keen perception of the velocities of separate molecules of the gas, is placed in charge of each door, with instructions to open it to all particles in the compartment A approaching it with *more* than a certain fixed velocity V, and only to such particles in A; but, on the other hand, to open it to all particles in the compartment B approaching it with less than a certain fixed velocity v, which is not greater than V, and only to such particles in B. Every molecule entering B will then carry with it more energy than a similar molecule entering A, so that the temperature of B will rise and that of A will fall, without any loss of heat or expenditure of energy. That is to say, by the application of intelligence alone, a mass of gas at uniform pressure and temperature may be divided into two parts, in which both the temperature and the pressure are higher in one than in the other, and from which work can therefore be obtained at the expense of heat. Assuming that this process could be continued indefinitely, by increasing V and decreasing v, without liquefying the gas, the whole of the energy in A would ultimately be transferred to B. It follows, therefore, that the principle of the dissipation of energy holds good only in virtue of the fact that we have no means of constraining the actions of individual molecules, although we can deal with their aggregates in bulk, owing to the extreme smallness and great number of the molecules, and the steadiness of their average behaviour. If we could deal with the molecules individually, we should be able to develop on a large scale what takes place continually on a very minute scale in any mass of gas—the occasional transient aggregation of warmer molecules in one small region, and of colder ones in another.

The principle that the energy of any material system, not intelligently controlled, tends to become mechanically disorganised, and never spontaneously tends towards organisation, applies, within the limits of our experience, to all the known phenomena of the material universe, and may be stated in the form that the unavailable energy of the material universe tends to a maximum, or, as Clausius has expressed it, its "*entropy*" (see Appendix I) tends to a maximum. It is shown, moreover, in Appendix K, that it is of mathematically infinite improbability that any universe in which matter is molecularly constituted should form a conservative system, except under the influence of intelligent directive power capable of counteracting the ten-

dency to the increase of entropy. This tendency can, as we have seen, be partially, but only very partially, counteracted by processes directed by human intelligence. The conclusion appears to be inevitable that if the universe contains, as some have maintained, no higher form of intelligence than that of man, it cannot be a self-contained conservative system. It must, under such circumstances, have been called into existence by some external agency, and must ultimately come to the end of its activity, viz. when the whole of its energy has passed into the unorganised form.

Some considerations, due to J. H. Jeans (see p. 295), show that this tendency of energy to become unorganised might conceivably be compensated if the universe contained portions of matter built up of electrons whose charges had a different value from that of the electrons in the matter with which we are familiar. If, for example, it were possible for the nature of the electrons to undergo some change, involving an alteration of their charges, in the processes incidental to the formation of nebulæ, as considered in Chapter XX., the changes might be of such amount and so distributed as to reorganise more or less completely the energy of the system which had become unorganised in other stages of its life. Such changes are certainly not more inconceivable than the formation of the electrons in the ether, but the argument set forth in Appendix K shows that they could produce complete compensation only under the action of power intelligently and purposely directed to the attainment of this object.

If two material systems in contact, or otherwise capable of mutual interaction, are such that there is no tendency to an exchange of energy between them, they are said to be in thermal equilibrium. They are then said to be at the same temperature, a statement which may be considered as a definition of the meaning of the term temperature. The mechanism of this interchange of energy has been elucidated in the case of gases, where it is effected by encounters between the molecules. There is therefore no tendency to such an interchange when the average energy of the bodily motions of the molecules is the same for each. For all states of matter the equilibrium of energy between systems capable of mutual interaction must necessarily involve the condition that a definite molecular state in one system shall be in equilibrium with a definite molecular state in the other. Larmor points out that the universality of this principle requires that this condition should be a very fundamental one, and argues as follows :—[1]

"It would seem that we can make at any rate an advance towards a

---

[1] *Loc. cit.*, p. 264.

complete view by realising that, even if our sensations of heat had not compelled us to assign a fundamental place to temperature in the physical scheme, the principle of negation of perpetual motions must have led to the formulation of that conception, just as it has, in fact, led to the conception of potentials. If thermal equilibrium between two homogeneous bodies A and B in contact were not conditioned merely by some physical property of A alone being equal to some property of B alone, then, if we had A in contact with B, and B with C, each in a state of equilibrium, and, removing B by mechanical means, moved A into direct contact with C, but with such ideal constraint applied to the matter in bulk that chemical action is prevented, the physical state of each of these latter bodies would become changed, involving the performance of mechanical work; and a self-acting cycle could be designed by which we might thus obtain an unlimited quantity of work, that is, so long as there remained any diffused molecular energy to be converted. Hence in equilibrium there must be a property, namely, the temperature, of each individual body in the field, that has the same value for all of them; although, of course, this does not prevent us from imagining a partition or constraint, nearly adiathermanous, across which such equilibrium would be established as slowly as we please. It follows also that equilibrium of temperature must be the same whether it is brought about by conduction or by radiation. Temperature, as thus introduced, has nothing to do directly with the field of force in which the body is situated; for the relations of bodies to fields of force, in which they are moved about, are treated independently in the consideration of energy relations, and must not be introduced twice over—or, in other words, the perpetual motion principle can be directly applied."

If the nature and measure of heat as a form of energy had never been experimentally determined, the law of the conservation of energy could not have been expressed in its most general and complete form, for we should have had no knowledge of what had become of the energy which had ceased to be mechanically available; so that, as Larmor points out, the principle forming the basis of general non-molecular physics is rather that of the limited conservation of available energy, as observed in completely reversible cycles or conservative systems, than the more general one of the complete conservation of the total energy.

Helmholtz has shown how the form of the function representing the available energy, except its undetermined part, can be experimentally determined for the different states of matter, and that the equilibrium state of any system of mutually reacting bodies at any assigned temperature is then the one which makes this function a minimum for that temperature, thereby formulating the general solution of the problem of physical and chemical equilibrium.

CHAPTER XI.

CONTACT ELECTRIFICATION AND ELECTROLYSIS.

WHEN two solid dielectrics are in contact along a surface, the superficial molecular layers will be within reach of each other's influence, and strong orientations of the polar molecules in these layers will be produced, accompanied by stresses across the intervening ether. The transmission will be effected partly by an intrinsic hydrostatic pressure, and partly by tangential elastic tractions arising from the ether twist. This ether twist, or electric displacement, in so far as it is not along the interface, will give rise to difference of electric potential between them.

Exactly similar considerations apply in the case of contact between two conductors, but the internal equilibrium of each conductor now requires that the potential shall be uniform throughout it, and therefore the surface stress system will adjust itself in such a manner as to make the potential difference constant over the interface. The result of this adjustment will be to direct the axis of the ethereal twist at every point of the interface more nearly along the common normal; that is to say, to make the electric displacement act along the normal, and so greatly diminish the potential difference which would otherwise be produced by the superficial layers. Much smaller difference of potential, therefore, and consequently much smaller superficial electrification, is produced by the contact of conductors than by that of dielectrics.

If two dielectric surfaces are rubbed together it is clear that the average distance between the two sets of molecules will be considerably diminished, and the mechanical disturbance will also, as in the case of magnetisation, assist the orientation of the polar molecules. For both reasons the potential differences, and consequently the superficial electrification, will be increased by the friction.

The contact phenomena between a solid and a fluid, whether liquid or gaseous, will differ from those which arise in the case of two solids; for, after a sufficient lapse of time, the mobility of

the fluid will permit of an adjustment of charged dissociated ions along its surface tending to ease off the internal stress, so that the boundary of the fluid will become completely, and more or less permanently, polarised.

Take, for example, a zinc and a copper plate immersed in dilute sulphuric acid, forming a simple voltaic cell. The electric stress in the interior of the solution will become null, and the potential difference between the two metals will be the difference of the potential differences between each of them and the acid. It will not, however, be the same as the potential difference due to direct contact between the metals. If the two metals are now connected by means of a wire, or other conductor, the potential difference maintained between the plates will cause a current round the circuit, and the energy required to maintain the current will be supplied by the solution of the zinc in the acid. In this case the current will soon begin to diminish, owing to what is known as electrolytic polarisation. As the zinc dissolves in the acid, forming zinc sulphate, hydrogen is liberated, and, adhering in small bubbles to the zinc plate, gives rise to a contact potential difference opposed to that which is driving the current round the circuit. Descriptions of the principal types of voltaic cells, and of the methods of eliminating electrolytic polarisation with more or less completeness, will be found in any elementary treatise on electricity.[1] If the zinc plate were perfectly homogeneous, no solution would take place when the circuit is incomplete. The ordinary zinc of commerce is, however, never completely homogeneous, so that local electric circuits are formed, giving rise to solution on open circuit, and also causing waste when in operation, as some of the energy liberated by the solution of the zinc is wasted in maintaining these local circuits. The local action may, however, be very nearly eliminated by carefully amalgamating the zinc plates with mercury, which makes their surfaces practically homogeneous.

Consider now a circuit consisting only of different metals, and known as a Volta's chain. In the absence of chemical action and differences of temperature, such a circuit must, according to the energy considerations discussed in the preceding chapter, be in equilibrium. No electric stress will be transmitted through any metallic link of the chain, but will be transmitted entirely by the ether surrounding its sides, and forming the interfacial layers between it and the adjacent links on either side. Moreover, the potential in the ether all round the surface of the same metal is uniform, and this uniformity applies to each link in the chain. The sum round the chain of the very rapid changes of potential

[1] See also W. R. Cooper's *Primary Batteries*.

which occur in crossing the different interfaces is therefore null, which is Volta's law of potential differences for metallic conductors.

Suppose now that the equilibrium is disturbed, by the introduction, for example, of a layer of an electrolyte at an interface. This will introduce a supply of chemical potential energy, which may be employed in maintaining a current round the circuit, so that the equilibrium will be disturbed. There will no longer be a uniformity in the potential surrounding each metal, and there will therefore be a change in the electric displacement, that is to say, in the state of ether twist, in the surrounding medium, which will diffuse into the metals, and give rise to a flow of electrons, which, in the final steady state, will be uniformly distributed throughout them.

Any theory of electrolysis must, in the first place, be in conformity with Faraday's two laws of electrolysis. If we call the mass of an electrolyte which is decomposed by the passage through it of the unit quantity of electricity its electrochemical equivalent, Faraday's laws may be stated as follows :—

1. The quantity of an electrolytic ion decomposed by a given current in a given time is equal to the product of its electrochemical equivalent by the current and by the time.

2. The electrochemical equivalent of any electrolytic ion is equal to the product of its chemical equivalent by the electrochemical equivalent of hydrogen.

Larmor expresses these two laws in the following single statement (*Æther and Matter*, p. 289), which is a convenient one from our present point of view :—

"The number of molecules of the anion liberated in any time is the same as the number of molecules of the cathion, and, corresponding to the liberation of one molecule of either of them, the same quantity of electricity passes in the current, a quantity which, on comparing different anions and cathions, is proportional to their chemical valencies; a definite quantity of electricity—the fundamental unit of charge, or the electron—thus corresponding to each valency, whatever be the electrolytic substance."

The valency of an element may be defined as the number of atoms of hydrogen which one atom of the element is capable of combining with or of replacing in a chemical compound. The same element may have different valencies under different conditions; for example, when oxygen combines with hydrogen under ordinary circumstances it forms water, H_2O, so that in this case oxygen is divalent. It may, however, under certain conditions, form hydroxyl, HO, in which case the oxygen is monovalent.

A very important advance in our knowledge of the phenomena of electrolysis was made by Hittorf, who, as the result of an

extensive series of investigations extending from 1853 to 1859, succeeded in demonstrating the existence of a certain ratio between the transport velocities of the cathion and the anion, which was independent of the potential difference between the electrodes, and, within certain limits, independent of the concentration of the solutions. At ordinary temperatures this ratio was not found to vary appreciably with the temperature, but tended towards unity when the temperature became considerable. During the course of his researches, Hittorf repeatedly insisted on the probability that a deeper insight into the nature of electrolysis might be obtained if the specific conductivities of various electrolytes could be measured and compared. At this time, however, no practicable method of making such measurements was known, owing to the difficulties arising from so-called polarisation effects.

Kohlrausch's discovery in 1880 that the Wheatstone bridge method of measuring resistances could be applied to the resistances of electrolytes, by the use of alternate currents and a telephone as indicator, in place of a continuous current and a galvanometer, made it possible to carry out the investigations suggested by Hittorf.

In order to obtain results which could be compared with one another, Kohlrausch varied the concentrations of the different solutions investigated in such a manner that each of them contained the same number of molecules per unit volume. In this manner he succeeded in obtaining the relative molecular conductivities, as he expressed it, of various electrolytes in solution. These measurements showed that the specific conductivity of an electrolyte decreased with decreasing concentration, but less rapidly, showing that the molecular conductivity increased as the concentration was diminished.

It was found by von Helmholtz in 1881 that this increase did not continue indefinitely as the solutions became extremely dilute, but attained a maximum value at a definite limit of dilution. It was found that this maximum value of the molecular conductivity was, in the case of electrolytes formed of two monovalent ions, simply the sum of the two migration velocities of the constituent ions. In the case of electrolytes containing polyvalent ions, the molecular conductivity was found to be somewhat smaller than this sum.

One of the first questions arising from the results of Faraday's investigations, as set forth in his two general laws of electrolysis, was the cause of the separation of the molecules of an electrolyte into its constituent ions, which always accompanies the passage of an electric current through it. Grotthus, in 1805, suggested

that this separation was due to the energy of the electric current, the atoms being in the first place oriented by the electric force at every point in the solution, arising from the potential difference maintained at the electrodes, forming what is often known as a "Grotthus Chain." It would be assumed, for example, on this theory, that in the electrolysis of hydrochloric acid, HCl, into hydrogen and chlorine, the atoms were first of all turned so that the hydrogen atom faced the cathode, while the chlorine atom faced the anode, and that the electric force then separated the molecules into their constituent atoms, enabling the positively charged hydrogen atoms, under the influence of the electric field, to travel towards the cathode, while the negatively charged chlorine atoms travelled towards the anode. This migration of the ions was supposed to take place by means of successive interchanges between adjacent molecules of the chain, until the last hydrogen atom was liberated at the cathode, and there gave up its charge, while the chlorine atom at the other extremity of the chain gave up its charge to the anode. For half a century this was accepted without question as the true explanation. Clausius, however, pointed out, in the year 1857, that it was not in accordance with the observed fact that electrolytic conduction begins as soon as the potential difference between the electrodes is sufficient to overcome the E.M.F. arising from the so-called polarisation set up by the passage of the current. If the explanation proposed by Grotthus were the true one, conduction should begin only when the potential difference exceeds the opposing E.M.F. due to polarisation by an amount sufficient to give rise to the separation into ions, and then there should be a sudden rush of current, in place of the gradual rise actually observed, as the potential difference between the electrodes is increased.

It was suggested by Svanté Arrhenius in 1884 that some of the molecules in an electrolytic solution already exist in the solution in a state of dissociation into their constituent ions, carrying positive and negative charges respectively, and that the effect of setting up a potential difference between the electrodes is simply to set these charges in motion, in obedience to the electric force set up within the electrolyte. According to the dissociation theory, as it is called, electrolytic conduction is due entirely to these free ions, the ions which are combined into molecules taking no part in the process. The ratio of the molecular conductivity of an electrolytic solution at any concentration to the sum of the migration velocities would, on this theory, represent the proportion of dissociated, or ionised, molecules present in the solution. This ratio has therefore been called the *ionisation coefficient*.

This theory gives a simple and satisfactory explanation of the elementary facts of electrolysis. For example, it accounts for the conductivity of fused electrolytes by the known action of heat in producing dissociation. The fact that many simple electrolytes, such, for example, as pure sulphuric acid entirely free from water, are non-conductors, is simply explained as being due to their molecules not being ionised.

Water is such a bad conductor, when purified to the highest degree possible, that it is usually considered that it would be a non-conductor if it could be obtained in a perfectly pure state. Water is therefore not supposed to take any part in the electrolysis of aqueous solutions, but to act merely as an ionising agent. It is unapproached by any other solvent in its power of causing dissociation. The lower conductivity observed by Kohlrausch, in the case of electrolytes with polyvalent ions, finds a simple explanation as being due to incomplete ionisation, even at the highest dilution at which measurements can be made.

The dissociation theory appears to introduce certain fresh difficulties in connection with some of the more recondite phenomena involved in the electrolysis of some complex chemical compounds,[1] and therefore it did not for a long time obtain general acceptance; but it has certainly proved to be a most convenient working hypothesis, in that it has made it possible to co-ordinate together various chemical phenomena which had hitherto not appeared to be related in any way to one another. It was only after the promulgation by Van't Hoff, in 1887, of his *osmotic* theory of solution, which would not at first sight appear to have any bearing on the theory of electrolysis, that the dissociation theory was generally accepted.

The existence of what is known as *osmotic pressure* in a solution is capable of very simple experimental demonstration. Let a glass cylinder, open at one end, be filled with a concentrated solution of sugar, and let the other end be then covered with a piece of bladder, in such a manner as to make an air-tight joint, and let the cylinder be immersed in a vessel containing water. The membrane will gradually swell outwards into a cup-like form, and if, after an interval of two or three hours, the cylinder is removed from the water, and the bladder pierced by a fine needle, a stream of liquid will issue from the hole. This phenomenon is due to the fact that the molecules of water are able to pass through the membrane into the cylinder, while the membrane is impervious to the sugar molecules. The result is an increase in the number of molecules within the cylinder, causing an increase of pressure, just as if they were molecules of

[1] See W. R. Cooper's *Primary Batteries*, pp. 113–116.

a gas. When a gas is contained in a closed vessel, we know from the kinetic theory of gases that the pressure on the interior of the vessel is due to the impacts of the molecules, and that its amount per unit of area is sensibly equal to a third of the mean value of the squares of the velocities of the gaseous molecules contained in unit volume of the gas multiplied by the density, provided the density of the gas is not sufficiently great for the result to be sensibly affected by the interactions of the molecules, that is to say, provided the gas is not approaching liquefaction. The experimental laws of Boyle and Gay Lussac, determining approximately the relations between the pressure, the volume, and the temperature of a gas, which are known to express the conditions more exactly the greater the rarefaction of the gas, can be shown to follow as necessary consequences of the kinetic theory when the rarefaction is sufficiently great to completely eliminate the results of molecular interactions. According to these laws, the product of the pressure and the volume of a given mass of any gas sufficiently far removed from the liquefying point, remains constant so long as the temperature is unchanged; and when the temperature varies, this product varies in direct proportion to the absolute temperature. Another well-known law, approximately obeyed by all gases, except when approaching liquefaction, is that of Avogadro, according to which, equal volumes of gases, measured under identical conditions of pressure and temperature, contain equal numbers of molecules.

It was pointed out by Van't Hoff, in the year 1887, that the results of observations made by de Vries, Pfeffer, and others on the osmotic pressure of solutions led to the conclusion that this pressure was the exact analogue of the pressure of a gas, and that the molecules of a substance in solution behaved, in relation to their volume, pressure, and temperature, exactly like the molecules of a gas, as shown by their being subject to these gaseous laws. He even found that the constants occurring in the expression of these relations for a dilute solution were the same as for a gas, leading to the definite law that the osmotic pressure exerted by the molecules of a given weight of substance in solution is equal to the pressure which an equal weight of the substance would exert at the same temperature, if it occupied a volume in the gaseous state equal to that of the solution. He therefore inferred that when a substance enters into solution it is thereby, to a very great extent, disintegrated into its separate molecules; and that when a sufficient degree of dilution is attained, the disintegration becomes complete, and the osmotic pressure then becomes a measure of the number of molecules present. This is Van't Hoff's theory of solutions.

The laws of osmotic pressure derived from diffusion observations

were subsequently confirmed by Raoult and others by quite different methods. It has been known for a long time that the vapour pressure of a liquid is lowered by dissolving another substance in it, and, as a consequence of this, the boiling point is raised, and the freezing point lowered. A series of investigations of this kind led to the following results:—Equimolecular solutions of any substances, that is to say, solutions containing equal numbers of molecules of the different substances when prepared with equal weights of the same solvent, and provided they are not electrolytes, give rise to equal osmotic pressures, equal relative depressions of vapour pressure, equal elevations of the boiling point, and equal depressions of the freezing point. In the case of electrolytic solutions it was found that the osmotic pressure, and with it the variations in the vapour pressure and the boiling and freezing points, were larger than was anticipated, the solutions behaving as if they contained an abnormally large number of molecules, and the effect was found to increase with the dilution, very dilute solutions being found to behave as if containing double the theoretical number of molecules. These results are completely accounted for by the dissociation or ionisation theory of Arrhenius, according to which very dilute solutions are completely ionised, so that the number of molecules is then doubled, and they may therefore be regarded as affording strong confirmation of that theory, provided that we can account for the osmotic pressure being dependent only on the number of foreign nuclei present in the solution, whether these are molecules or merely ions. This we shall presently do.

We saw at the beginning of this chapter that contact between two solids, or between a solid and a fluid, gives rise to a difference of potential between them. Small differences of potential are also set up when different electrolytes are brought into contact, even when they consist of similar materials, and differ only in concentration. Here, however, the exciting cause cannot consist in strains in the surrounding ether. Nernst explains it as being wholly due to the differing velocities of the ions which are set in motion by osmotic pressure, the result of which would be to produce an excess of cathions in one solution and of anions in another, giving rise to a positive electrification on one side of the transition layer, and a negative electrification on the other side. Such an arrangement is called a *liquid cell*. The E.M.F.'s of such cells are very minute, and quite insufficient to overcome the electrostatic equilibrium which is very soon established between the cathions and the anions. The current will therefore continue for a short period only. When a cell is formed by two solutions of the same metallic salt in the same solvent, differing, therefore,

only in concentration, and the electrodes are of the same metal as the one in solution, the cell is called a *concentration cell.* The electrode in the more concentrated solution becomes positively charged by means of the positive charges given up to it by the cathions as they pass into the metallic state, while at the same time fresh cathions are formed by the solution of the electrode in the solution of lower concentration which receives the negative charges. It will be seen that this process tends to equalise the concentrations of the two solutions, and when this is accomplished the difference of potential will disappear, so that the cell will cease to produce a current.

The osmotic pressure between a solution and the pure solvent is really nothing else than the mean aggregate of the forces which must be applied to the individual molecules to prevent them from travelling across the interface into the pure solvent, whether applied by the resistance of a material partition or, as in the case of the ions diffusing across the interface separating two electrolytes in contact, by the pull of the electric field arising from the diffusion, the unmodified molecules of the solvent being, in each case, free to move either way.

Sir Joseph Larmor has shown that this may be directly deduced from the principles considered in the preceding chapter. Suppose the pure solvent and the solution, such as water and sugar solution, to be separated by a rigid porous partition, and suppose an external hydrostatic pressure to be applied to the whole surface of the partition on the side of the sugar solution, of just sufficient amount to prevent transpiration. Now suppose this pressure to be slightly reduced, so as to allow of a small amount of transpiration. This transpiration of the water into the sugar solution will involve work against the applied pressure, which will be measured by the product of the pressure into the volume of water so introduced into the solution. If the operation is allowed to continue so slowly that it always takes place under conditions of equilibrium, it may be reversed by simply reversing the pressure, so that the work is done at the expense of an equivalent of the *available* energy, no dissipation of energy coming into play. Part of this will be thermal energy, and part will be energy of a molecular type which would otherwise have been degraded into heat owing to the mixing of the liquids. The osmotic pressure between the two liquids must therefore be measured by the whole amount of *available* energy that would be so degraded into heat when unit volume

of the water is mixed with an indefinitely great volume of the solution into which it transpires, assuming that there is no sensible change of volume during the process. If there is a change of volume, then this value must be altered in the ratio of the final volume to the initial volume of the water which has transpired. If there is any such change of volume, it must be independent of the concentration, provided that the molecules are too far apart to influence each other. The argument will therefore be exact if the pores in the partition are so narrow that the cross sections of the filaments of water contained in them each involve so few molecules that the mutual energy of the molecules of water in the pores is negligible in comparison with that of an equal mass of the water in bulk. Since osmotic pressure arises only when the tubes or pores in the partition are of molecular fineness, it follows that it is not an ordinary transmitted mechanical pressure. The energy which is associated with it is, therefore, not merely the *organised* energy from which the mechanical forcive is derived, but the whole amount of *available* energy, *organised* and *unorganised*. In the case of wider pores or tubes traversing the partition, the mutual energy of the molecules in them will no longer be negligible, so diffusion will take place within the pores, involving a dissipation of energy which would otherwise have added to the osmotic pressure, and the latter will therefore be reduced in amount.

Let us suppose that the pressures on opposite sides of a porous partition separating two dielectric fluids are adjusted so that there is no transpiration, and that an electric field is then introduced. The result of this will be that the equilibrium will be disturbed by the electric tractions on the interfaces between the two fluids in the interior of the pores, and in order to re-establish it, it will be necessary to introduce a compensating pressure. That is to say, an electric field is capable of altering the osmotic pressure between two dielectric fluids. This electric effect will continue to act, even in the case of comparatively large pores, and even in the case of liquids which mix readily, for the electric tractions hold good for gradual transitions, and the obliteration of the sharpness of the interfaces within the pores will proceed extremely slowly. If a stream of conducting liquid is forced through a porous dielectric partition, it will give rise to an electric current across the partition.

Conversely, an electric current made to traverse the partition will carry some of the liquid with it. For, over the surface of each pore there will be a difference of potential between the partition and the liquid, arising, as explained at the beginning of the present chapter, from the polar molecules lying along the interface becoming strongly oriented owing to their mutual actions. This difference of potential will in time be reduced, owing to free ions from the liquid being attracted and retained, giving rise to an opposing potential difference, forming the origin of what are known as *electrocapillary* phenomena. The observed current across the partition is due to some of these ions being carried along with the liquid. Larmor observes that a sudden diminution in the extent of surface would act similarly by crushing out the ions, as in the observed electrification near waterfalls and other places where fine spray is formed. Rapid extension of the surface, on the other hand, as, for example, in the formation of drops in air, should eliminate the effect of the ions held by the polarised air-film on the surface, by spreading them over a wider area, and so increase the potential difference towards its maximum value, viz. that due to contact between the water and the air, in the absence of electrons.

In very dilute solutions the osmotic pressure will tend to approach a limiting maximum value, for, after a certain stage of dilution has been reached, the dissolved molecules will be effectively outside one another's range of action, and therefore each one of them will be influenced only by the molecules of the solvent which completely surround it. Further dilution will, therefore, merely cause a wider separation between the dissolved molecules, with such energy changes as may be involved, and will not give rise to sensible change in the mutual available energy of the molecules of the solvent and the solute respectively.

Let us suppose, in the first place, that the solute or dissolved substance is a gas, and that the solution is separated from the pure solvent by a partition which the solvent is capable, and the gas incapable, of traversing. The solvent will then transpire across the partition into the solution, unless opposed by a definite osmotic pressure of sufficient amount to produce equilibrium. Now suppose that a certain volume of the solvent is allowed to transpire through the partition into the solution, the transpiration taking place extremely

slowly, so as to avoid any dissipation of energy. The only essential change in the system will be the expansion of the collection of molecules of the contained gas, each having the same fluid environment as before, into a larger space. We may, therefore, compare directly the available energies of the two masses of gas, and suppose that the fluid environment has been directly accomplished in each case by means of some ideal process not involving any expenditure of energy. The change in available energy due to the dilution will therefore be simply that which corresponds to the free gaseous expansion of the dissolved gas. Thus we have obtained a theoretical demonstration of Van't Hoff's law that the osmotic pressure of a very dilute solution is equal to the gaseous pressure of the dissolved molecules *when they are supposed to occupy the same volume as in the gaseous state.* In the case of dissolved liquids and solids we cannot actually have the molecules existing free at the same volume as they occupy in the dilute solution, but, as Larmor points out, the reasoning may legitimately be extended to them on account of the following considerations:—

The circumstance which makes the above purely imaginary process legitimate is that the available energy is a function of the constitution of the matter *in bulk*, not depending on the accidental characteristics of state or motion of the individual molecules. Now, the only change that has occurred as regards the constitution of the substance in bulk, that can affect either the available or the total energy, is the change of volume of the solution by transpiration of the pure solvent across the partition, which, by the above, affects it in a manner absolutely independent of the nature of the homogeneous solvent, and therefore of the existence of the solvent at all, because the relation of each molecule to the portion of the solvent within its sphere of influence is not changed. Thus the final result arrived at is that the osmotic pressure depends only on the number of distinct particles, or nuclei, and on no other circumstances.

The expression obtained experimentally by Raoult for the relative depression of vapour pressures can be theoretically deduced from the law of osmotic pressure, the theoretical validity of which has now been demonstrated, and the experimental determinations of the elevations of the boiling point, and depressions of the freezing point, are also in close accord with the theoretically

derived expressions. It therefore becomes possible to determine osmotic pressures and relative depressions of vapour pressure, which are difficult to observe with accuracy, from the easily made observations of depression of the freezing point.[1]

The investigations of Hittorf and Kohlrausch, referred to earlier in the chapter, suggest that the mobilities of the anion and the cathion are mutually independent, and in 1886 Sir Oliver Lodge described a method of confirming this by making the motions of the ions visually perceptible. Each must, therefore, carry, as an electric charge, a number of electrons proportional to its valency. There must be an equal number of anions and cathions in every element of volume, to a very close approximation, for the number of free ions per unit of volume is very large, so that a very small relative difference in the numbers of positive and negative ions would give rise to sensible electrification of the solution. The independent transport velocities of the ions must be regarded as due to the electric force acting on them at every point in the solution, this being equal in amount to the electromotive force per unit of length. The result will be that the ions with the greater speed of transport will reach the electrode towards which they are travelling in greater numbers than the ions with the lower transport speed. Now, according to Faraday's law, the ions are liberated in equal numbers at each electrode, so that, as the electrolytic process continues, there will be a tendency to an excess of the more mobile ions in the neighbourhood of one electrode, accompanied by a deficiency of the less mobile ions in the neighbourhood of the other electrode, thus leading to changes of concentration such as were observed by Hittorf. In order that the electrolytic current may remain steady, this increase of density in the electrolyte in the vicinity of one of the electrodes, and decrease in the vicinity of the other one, must be relieved in some way.

Sir Joseph Larmor shows how the method of relief may be very simply arrived at. Consider the beginning of the electrolysis in a fresh uniform solution, and let v_1 be the average speed of drift of the cathion to the right, and v_2 that of the anion to the left. Suppose the speeds of the respective ions to be represented as follows:—

$$\begin{aligned}\text{speed of cathion to the right} &= v_1 = \tfrac{1}{2}(v_1 + v_2) + \tfrac{1}{2}(v_1 - v_2);\\ \text{speed of anion to the left} &= v_2 = \tfrac{1}{2}(v_2 + v_1) + \tfrac{1}{2}(v_2 - v_1).\end{aligned}$$

The motions of the ions may therefore be considered as consisting of—

(1) A drift of the two ions right and left respectively, with the speed $\frac{1}{2}(v_1 + v_2)$.

[1] See Poynting and Thomson's *Heat*, chapter xix.

(2) A drift of the two ions together to the right with the speed $\frac{1}{2}(v_1 - v_2)$.

(1) represents a conduction current involving no accumulation of ions at either electrode; while (2) represents a uniform flow of the electrolyte without any electric separation, and leads to an increase of density in the vicinity of the cathode, or of the anode, according to whether v_1 is greater or less than v_2.

At the beginning of the process the concentration near the centre will be sensibly constant, so that v_1 and v_2 will then be proportional to the mobilities of the two ions respectively, so that

$$\frac{\text{total mass transported across the centre}}{\text{electrolysed mass transported across the centre}} = \frac{v_1 - v_2}{v_1 + v_2};$$

and hence

$$\frac{\text{amount of cathion transported}}{\text{amount of cathion electrolysed}} = \frac{v_1}{\frac{1}{2}(v_1 + v_2)};$$

$$\frac{\text{amount of anion transported}}{\text{amount of anion electrolysed}} = \frac{v_2}{\frac{1}{2}(v_1 + v_2)};$$

which are the transport numbers experimentally determined by Hittorf.

It is shown in Appendix L that when steady diffusion is allowed to go on, the back electromotive force at the junction of the fluids is thereby reduced in the ratio of the difference of the ionic speeds to their sum.

Now let us suppose that an extraneous magnetic field H is established transverse to the electrolytic current. Let e be the charge carried by each ion, let v_1 and v_2 be their velocities of drift under unit electric force, and let E be the electric force, at each point, driving the current, that is to say, the E.M.F. per unit of length. Then the forces acting on the individual ions will be ev_1EH and ev_2EH, both acting in the same direction, and tending to drive the electrolyte as a whole sideways. They will also tend to heap up the two sets of ions sideways at different rates, and thus produce electrical separation, resulting in a differential transverse electric force, which can be shown to be uniform and independent of the concentration, so that it arises from a purely superficial electrification on the sides of the electrolyte. This is known as the Hall effect, its existence having been suspected by him from general considerations, although he did not express it in as definite a form as is now possible in accordance with the electron theory. He also succeeded in experimentally demonstrating its existence in metals.[1]

[1] See Chapter XIV., p. 318.

The question might be asked whether electric separation in an electrolyte might not be produced by an imposed magnetic force at right angles to the direction of the earth's motion. The general equations of electrolytic conduction, derived from the considerations in this chapter, show, however, that when the electrolyte is not traversed by an electric current, no electric force will be produced within it from its motion through the ether, as was found to be the case in a metallic conductor; the reason in the case of the electrolyte is that a Hall effect is produced which counterbalances the induced electrostatic field arising from the convection. When the electrolyte is traversed by an electric current, the Hall effect will be diminished by this electrostatic field, but, since the field contributes nothing to the electromotive force round the circuit, there will be no alteration that could be detected with a galvanometer.

Electrolytic conduction is affected, to a minute and quite imperceptible extent, by the earth's motion. If this convective velocity is v, and is in the direction of the current, the velocities of the positive and negative ions relatively to the moving electrolyte will be

$$\frac{v_1}{1+\frac{vv_1}{c^2}} \text{ and } \frac{v_2}{1-\frac{vv_2}{c^2}},$$

assuming the action between the ions to be entirely electric. The electric current, being determined by the sum of these velocities, will therefore be diminished in the ratio

$$\frac{v_1+v_2-\frac{v}{c^2}(v_1{}^2-v_2{}^2)}{v_1+v_2}=1-\frac{v}{c^2}(v_1-v_2).$$

The conductivity will therefore be diminished in the same ratio, where v_1-v_2 must now be taken as the *average value* of the difference of the velocities of drift of the positive and negative ions. It will be noted that the effect is reversed on reversing the direction of the current relatively to the motion of the electrolyte through the ether. It is at most only of the second order of small quantities, and if the velocities of the positive and negative ions were equal, it would be reduced to a fourth order effect.

CHAPTER XII.

OPTICAL PHENOMENA.

THE ordinary elementary laws for the reflection and refraction of rays of light incident on an interface separating two isotropic media, assumed to be perfectly transparent, require no assumptions as to the nature of the media beyond what is involved in the condition that there shall be no loss of energy in the act of reflection or refraction. There must, therefore, be no energy carried away by longitudinal waves, a condition which is necessarily fulfilled by material media of any character permeated by ether of the type we have considered, possessing purely rotational elasticity. For we have seen that electromagnetic waves in an ether of this type will give rise to no longitudinal waves, even should the ether possess finite compressibility. We must bear in mind, however, that no material medium is perfectly transparent, so that laws established on the assumption of perfect transparency can only be regarded as approximations. In order that a ray traversing obliquely an interface between two such media may undergo refraction, it is merely necessary that the velocity of transmission should have different values in the two media. Reflection of a ray falling on an interface will take place whenever there is a nearly abrupt transition between two media, so that the surface waves penetrate to a depth which is either very large or very small in comparison with the thickness of the transition layer. Under these circumstances the solution of any problem in reflection or refraction is obtainable by the purely geometrical methods applicable to systems of rays, the rays being most simply defined, as in Chapter VIII., as the paths of the energy. These approximate laws can also be derived as necessary consequences of the electromagnetic boundary conditions which must now be considered, independently of the form which these conditions may assume in special cases.

At an interface separating two media of different dielectric qualities there will be a discontinuity in the electric elasticity as we pass from one to the other, and a consequent discontinuity in

the normal component of the electric displacement. There will be a resulting tangential electric stress in the ether, which must necessarily be continuous. Now, this is equal to the tangential component of the electric force, but at right angles to its direction. The tangential components of the electric force must therefore be continuous.

Again, in a magnetic medium the magnetic induction represents the smoothed-out velocity of flow due, partly to motions communicated by distant disturbances, and partly to the magnetic vortices of the medium. The magnetic induction is therefore always a stream, and its normal component must therefore be continuous at an interface. With respect to the tangential component, however, the whole induction must be divided into two parts, one representing the effect of the neighbouring vortices, and the other including the rest of the flow. The latter portion is what is usually known as the magnetic force, and is clearly the only part whose tangential component must be continuous, when crossing an interface separating media with different magnetic properties.

We see then that the tangential components both of electric and magnetic force are continuous when crossing an interface between two media, and these are the necessary and sufficient conditions for the propagation of an electromagnetic wave.

This is easily demonstrated mathematically from the general equation representing the propagation of a plane electromagnetic wave, on writing down the relations between the electric forces involved in the boundary conditions. If the axis of z is taken in the direction of propagation, and the point of incidence on the interface as origin, then the components of these forces must, for $z=0$, all be proportional to a function involving only the time and the co-ordinates x and y. The forces in question may then be expressed in terms of the angles of incidence, reflection, and refraction, and when so expressed, the resulting equations give immediately the following laws:—

(1) The incident, reflected, and refracted rays all lie in the same plane, which is perpendicular to the interface.

(2) The angle of reflection is equal to the angle of incidence.

(3) The sine of the angle of incidence bears a constant ratio to the sine of the angle of refraction, for any given pair of media.

These are the well-known laws of reflection and refraction, and the constant ratio defined in law (3) is the quantity known as the index of refraction from the first medium to the second.

In order to extend these approximate laws so as to cover the case of crystals, also assumed to be perfectly transparent, it must be borne in mind that their magnetic permeability, as is the case for all dielectrics, is sensibly equal to unity. Their electric

properties will therefore be determined by the dielectric constant only, and, as these properties are different in different directions, this must vary with the direction of the electric force. It will then follow, as was pointed out in Chapter VIII., that the electric induction is not, in general, in the same direction as the electric force.

The diminution in the velocity of a ray of light passing from the ether into a material medium, which is the cause of its refraction, is, from our present point of view, due to the reduction in the effective elasticity of the ether arising from the presence of material molecules composed of electrons, and therefore capable of polarisation under the action of the changes in electric displacement constituting the electromagnetic wave-train which we call a ray of light.

Lorentz's approximate relation (see p. 132) between the dielectric constant K and the density ρ of a material medium, viz.,

$$\frac{K-1}{K-2}=A\rho,$$

shows that K and ρ increase together, and therefore accounts for the increase in refractive power which accompanies increase in the density of a given substance, as a result of the corresponding decrease in the speed of radiation.

A perfectly transparent body would have to be a perfect dielectric, as otherwise electric currents would be produced when it is traversed by electromagnetic waves, and the energy dissipated by them in the form of heat must be supplied by the radiation, so that absorption will take place. This absorption is of the kind that can be represented mathematically by the introduction of a "Dissipation Function" (see p. 51) into the expression for the energy from which the equations are derived. This must be the case if a radiation maintains its period of vibration unaltered in passing through absorbing, or viscous, media, and if this condition were not fulfilled, a strong beam of homogeneous light, after passing though a film of metal or other absorbing medium, would emerge as a mixture of light of different colours. The only case where anything of this kind has been observed is in the phenomena of fluorescence, so that, with the exception of these phenomena, the laws of absorption must be such as to be derivable from a single dissipation function. Now, the optical properties of iron and the other ferromagnetic bodies show that, even in their case, the magnetic permeability is sensibly equal to unity for the number of reversals of magnetisation per second involved in luminous waves. As a consequence of this fact, in combination with the derivability of the laws of absorption from a dissipation function, it

follows that the solutions of problems involving reflection and refraction in absorbing media can be derived from the solutions obtained for transparent media by replacing the real dielectric constants by complex ones, that is to say, by quantities of the form $P + Q\sqrt{-1}$, and separating out the real from the imaginary portions in the resulting formulæ. No further assumptions are required in order to account for the complicated phenomena of metallic reflection.

Refraction is always accompanied by *dispersion*, that is to say, the light of shorter wave-length is more refracted than that of longer wave-length, so that when ordinary sunlight is refracted through a prism, for example, it is spread out into a spectrum of which the visible portion begins with the less refrangible red rays and ends with the more refrangible violet rays. The ratio of the total dispersion to the mean value of the refraction varies greatly for different media, a fact which is utilised in the manufacture of so-called *achromatic* lenses, which are built up of two or more lenses formed of glass with different dispersive powers. It is never possible to eliminate the dispersion entirely and produce perfect achromatism, but the coloration is reduced to an almost insensible amount in some of the lenses of Jena glass. In some substances the dispersion is abnormally great at certain portions of the spectrum, resulting in a displacement of some of the colours from their normal relative positions, a phenomenon known as *anomalous dispersion.*

Ordinary, or normal, dispersion might arise from a discrete molecular structure in the medium, comparable in linear dimensions to the wave-lengths of luminous vibrations, but the actual number of molecules in a wave-length is too great to account for the whole of the dispersion. The remainder is, however, fully accounted for, on the electron theory, by the sympathetic vibrations excited in the groups of electrons forming the molecules of the medium by the periodic changes in electric displacement constituting the luminous waves. It is only necessary to make the assumption that the molecules have natural, or intrinsic, vibration periods of their own, capable of responding to these periodic impulses; and this is justified by the definite character of the spectrum obtained from the light emitted by molecules of a definite type when in the gaseous state. The origin of the periodicity of refracted rays emitted from incandescent solids and liquids will be considered presently. This periodicity is obviously essential to the validity of the explanation of dispersion as arising from sympathetic molecular vibrations. Anomalous dispersion will occur whenever the period of the luminous, or other electromagnetic wave, is nearly equal to one

of the intrinsic molecular ones. In all other cases the dispersion will be normal.

If electromagnetic waves were to impinge upon a perfectly conducting substance there would be no refracted wave; for such a substance, having no electric elasticity, would be incapable of transmitting the disturbance, as the surface electrons would move, under the action of the electric forces of the waves, without encountering any resistance, in such a manner as to neutralise them. The waves would therefore be entirely reflected without any loss of energy.

If, on the other hand, the waves were to impinge upon a perfect dielectric, these would disturb the Faraday tubes in the medium, including the electron pairs forming the material molecules, and set them in vibration. In the transition layer forming the interface, some of these vibrations would be travelling forwards and some backwards; and the former would give rise to the refracted waves, the later to the reflected waves. If the periods of vibration of any tubes capable of independent vibration, or of any of the groups of electron pairs forming the material molecules, corresponded to any of the periods of the incident disturbance, their vibrations would increase in amplitude, and tend to neutralise the electric forces due to the disturbance, so that the waves of these particular periods would be partly absorbed and partly reflected, but if the medium were of sufficient thickness, no waves of such periods would pass through unabsorbed. In the case of luminous waves, this gives rise to the phenomenon known as the selective absorption of light. We see from this that anomalous dispersion and selective absorption must always go together, as they arise from the same cause.

If the periods of the disturbance were sufficiently long as compared with the intrinsic periods of the molecules, there would be neither absorption nor dispersion, for the static portion of the dispersion, arising from the discreteness of the medium, also decreases as the period, and consequently the wave-length of the disturbance, increases.

As no actual material media are either perfectly conducting or perfectly non-conducting, their properties will be compounded of those which would belong to a perfect conductor and those which would belong to a perfect dielectric, with the addition of absorption due to viscous resistance to the movement of the free electrons.

Since the wave-lengths of light lie between three ten-thousandths and eight ten-thousandths of a millimetre, while electric waves are obtainable having wave-lengths varying from a few millimetres up to a couple of thousand miles or more, it will be understood how

it is that electric waves extending over a very considerable range are incapable of exciting sympathetic vibrations in material molecules, and can therefore be reflected by metallic conductors without sensible absorption, almost, that is to say, as if the metals were perfect conductors.

We saw in Chapter IV. that polarisation of luminous rays can be detected by the fact that the polarised rays have different properties in different azimuths in a plane perpendicular to the direction of propagation, and considered the propagation through the ether of a special kind of polarised wave, viz. a plane polarised wave, and this was assumed to consist of a vibration of very simple type, which was defined as a simple harmonic vibration.

If a pendulum consisting of a weight suspended from a long string is displaced through a very small angle, it will execute a series of vibrations which are sensibly of this character; that is to say, if the line of the string were produced so as to intersect a horizontal sheet of paper placed below the weight at a point P, P would describe a series of simple harmonic vibrations. The reason of this is that, as long as the angle through which the pendulum is displaced is small, the restoring force, due to gravity acting on the weight, is sensibly proportional to its linear displacement, that is to say, to its horizontal distance from the vertical through its position of equilibrium, which is also the vertical through the point of suspension.

If a particle in the interior of an isotropic elastic solid were displaced to a small extent, it would also execute vibrations about its position of equilibrium; and as the restoring force acting on the particle in any position, due to the elasticity of the substance, would be proportional to its displacement from the position of equilibrium, these vibrations would again be sensibly simple harmonic. If the body were not isotropic, but had different elasticities in different directions, the restoring force for small displacements would still be proportional to the displacement, but it would no longer be directed towards the position of equilibrium, that is to say, in the same direction as the displacement. This restoring force acting on the particle in any position could be represented by its components resolved in any two directions at right angles, lying in the plane of its motion, and these components would be proportional to the distances of the particle from two axes at right angles to each other, and passing through the equilibrium position. The magnitude and direction of the velocity of the particle at any point of its path would be obtained by compounding, by the law for the addition of vector quantities, the velocities due to the forces separately. It would therefore move in a path obtained by compounding two simple

harmonic motions at right angles. This path would therefore in general be an ellipse, becoming a circle if the two rectangular components were equal in magnitude, and the ellipse would be described harmonically.

In the case of the rotationally elastic ether, we saw in Chapter IV. that the ether displacements representing magnetic induction are simply the resultants of the twists of the elements of volume constituting the electric displacement, so that the displacement is not an actual displacement of a definite particle of ether corresponding to the displacement of a vibrating particle of an elastic solid. Variations in the electric forces, similar in direction and magnitude to the mechanical ones occurring in an elastic body, give rise to corresponding resultant ether displacements; and if the displacement at each instant of a vibration corresponds in magnitude and direction with that due to the movement of a vibratory particle in an ellipse, in the case of a vibrating elastic solid, we may still call the vibration an elliptic one.

The most general kind of polarised ray is one in which the vibration is elliptic, with circular polarisation as a special case. An elliptically polarised ray could be distinguished optically from an unpolarised one, but a circular one, as such, could not. However, it is possible, by optical apparatus, to actually resolve the circular polarisation into two plane polarised rays, with the planes of polarisation at right angles, and then to separate out one from the other. That is to say, we can physically imitate the mathematical resolution of a uniform circular motion into two simple harmonic motions at right angles to each other.

A motion might be imagined in which the vibration of a luminous ray at any one instant was of an elliptic character, but that the ellipse, instead of retaining its orientation constant, with its longest and shortest axes (its principal axes) always pointing in the same direction, continually changed its orientation. Such a vibration would not constitute an elliptically polarised ray.

If the motion round the ellipse, or circle, is clockwise to an observer looking in the direction of the source from which the ray is emitted, the polarisation is said to be right-handed, and left-handed when the motion is counter-clockwise to a similarly situated observer.

We have seen that a uniform circular motion is equivalent to, or can be decomposed into, two simple harmonic motions at right angles, and having the same amplitude as the uniform circular motion.

It can easily be seen that two uniform circular motions in opposite directions, and having the same amplitude, are equivalent to a single simple harmonic motion.

Consider a point P moving uniformly in a circle of radius $a = \text{OA}$ (fig. 20). Let $\text{OM} = x$ be the projection of OP on the diameter AB, and let $\text{ON} = y$ be its projection on the diameter CD at right angles to AB. Let the angle $\text{POA} = \phi$, and let t be the time when the point occupies the position P, let τ be the period, or time taken to complete a whole revolution, and let $t = \theta$ when P is at A.

Then we have

$$x = a \cos \phi = a \cos \frac{2\pi}{\tau}(t - \theta),$$

and

$$y = a \sin \phi = a \sin \frac{2\pi}{\tau}(t - \theta),$$

which give algebraical expressions for the simple harmonic motions with amplitude a and period τ along AB and CD respectively.

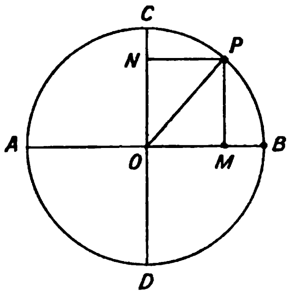


FIG. 20.

It is easy to show, by means of these expressions, that any number of simple harmonic motions in a plane, having the same period, but differing in amplitude and phase, may be combined into an elliptic motion, or into two uniform circular motions in opposite directions, but not necessarily along circles of equal radii.

Fourier has shown that any periodic motion of period τ may be decomposed into either a finite or an infinite series of simple harmonic motions having periods of

$$\frac{\tau}{1}, \frac{\tau}{2}, \frac{\tau}{3}, \frac{\tau}{4}, \frac{\tau}{5}, \text{ etc.}$$

The general theorem of which this is a case is not confined to periodic motions, its precise statement being as follows :—

Any function $f(t)$, however complex, provided it has only a finite range of singularities within the range considered, can be resolved, in the interval between the values $-\tau$ and $+\tau$ of the

variable t, into a series of simple harmonic functions of multiples of $\frac{\pi t}{\tau}$, there being two of these functions for each multiple r, differing in phase by a quarter period, viz. $\cos r\frac{\pi t}{\tau}$ and $\sin r\frac{\pi t}{\tau}$, so that

$$f(t) = a_0 + a_1 \cos\frac{\pi t}{\tau} + a_2 \cos 2\frac{\pi t}{\tau} + a_3 \cos 3\frac{\pi t}{\tau} + \text{etc.},$$

$$+ b_1 \sin\frac{\pi t}{\tau} + b_2 \sin 2\frac{\pi t}{\tau} + b_3 \sin 3\frac{\pi t}{\tau} + \text{etc.}$$

We shall see presently that, just as we can physically decompose an elliptic or circular vibration into the corresponding simple harmonic vibrations by optical methods, so the passage of more complex vibrations through various kinds of optical apparatus leads to their physical resolution into simple harmonic constituents corresponding to the mathematical resolution effected by Fourier's theorem, and it is for this reason that we are able to consider the simpler types of waves in accounting theoretically for optical phenomena.

Waves of simple harmonic type are propagated through free ether with a velocity independent of the wave-length, and therefore without change of type. Any plane wave may therefore be so propagated. A wave is said to be homogeneous when the displacements at *every point* are simple harmonic. Such a wave, travelling in the direction of the axis of z, that is to say, at right angles to the plane of xy, may be represented by the expression

$$x = a \cos 2\pi\left(\frac{x}{\tau} - \frac{z}{\lambda}\right),$$

where there is no limitation to the distance x. Its period will be τ, and its wave-length λ. Such a wave is, however, strictly speaking, a theoretical abstraction, for the equation must hold for all values of the time, as well as for all distances along the line of propagation. This could be the case, for waves emitted by a vibrating particle, only after the particle had been vibrating for an infinite time.

There is, in fact, physically no such thing as homogeneous light. A ray of the most homogeneous light obtainable by allowing light from an extremely narrow slit to pass through a series of prisms will in general show no sign of polarisation, and from the definition of a homogeneous wave, and of a polarised wave, it follows that all homogeneous light must be polarised.

If we apply Fourier's theorem to resolve any disturbance, we

may obtain an infinite series of terms of the same type as those which have been shown to represent simple harmonic motions, whereas mathematically the equation holds only for a certain time τ. We may, however, by taking τ large enough, express the whole disturbance as being formed by the superposition of a number of disturbances each of which may be made as nearly identical as we please with homogeneous light. Now, Gouy showed in 1886 that the proof, by the Fourier analysis, of the independence of the separate vibrations as regards energy, applies only to the average energy over a long range of time, and that this is relevant to optical investigations, because we may there treat the sources of light as being constant, which makes the average energy independent of the length of the time interval. White light may therefore be treated as consisting of a superposition of a number of periods differing very slightly from one another. These considerations, which were confirmed by Lord Rayleigh by means of detailed analysis in 1899, justify us in simplifying theoretical investigations by dealing with theoretically perfect homogeneous light as applied to the light emitted by vibrating gaseous particles which, from the definiteness of their spectra, clearly emit light which differs very little from theoretically homogeneous light over a very considerable number of vibrations.

These considerations do not, however, explain why the formless mass of radiation emitted from an incandescent solid or liquid should possess such periodicity as is required to account for the phenomena of dispersion and absorption. The explanation given by both Gouy and Lord Rayleigh is that the periodicity, in sunlight for example, required to account for these and other optical phenomena, such as colour sensation, is only acquired after passage through the optical apparatus employed in the observations, including the eye itself. Periodicity is not necessary in order to maintain the permanence of type of luminous waves traversing the isotropic ether, although it is necessary to permanence when the medium is ælotropic, and therefore we have no evidence of the existence of periodicity in the light prior to its being received by our optical apparatus.

A consideration of the application of the Fourier analysis to a single longitudinal pulse constituting a sound wave in an organ pipe will assist us in attaining a clear idea of the meaning of the mathematical resolution. Assuming this pulse to be of complex type, the Fourier analysis will, at any instant, resolve such a pulse into an infinite series of simple harmonic pulses of all possible wave-lengths, and each filling the whole tube, imagined as extending to an infinite distance in both directions. Each one

of these components forms a simple onward travelling wave-train of unlimited length, and if it existed alone it would be recognised as such by the ear or other suitable acoustic receiver. This almost appears to lead to the impossible conclusion that such a limited single pulse involves an infinite series of regular waves travelling both in front of it and behind it, although in reality there is no disturbance at all of the air in these positions. The explanation obviously is that this analysis is objectively valid only in the sense that it analyses the sequence of disturbances at the place occupied by the ear or other acoustic receiver, so that it proves that the effects there produced are the same as would arise from the infinite wave-train which, in itself, has no objective existence apart from the receiving apparatus.

This explanation is confirmed by the phenomena of polarisation, and is most simply illustrated by considering what happens when a ray of ordinary white light, such as sunlight, falls upon a doubly refracting crystal. The refracted portion of the ray is then split into two distinct rays. Within the crystal the elasticity of the ether is different in different directions, and consequently the velocity also varies with the direction, for, in the case of an isotropic medium, it can be shown that if v is the velocity of propagation, ρ the density of the medium, and R the elastic resistance to distortion, then

$$v = \sqrt{\frac{R}{\rho}}.$$

In an ælotropic, or doubly refracting, crystal the velocity will therefore vary with the direction, and it might at first sight be imagined that the entering ray could be scattered in all directions. What actually happens, however, is that the ray divides into two portions only, one following the path of least elasticity, and therefore of least velocity, and the other that of greatest elasticity, and consequently greatest velocity. Both of these rays become, in general, plane polarised, the planes of polarisation not being necessarily at right angles to each other, though this would be the case if there were no absorption. That is to say, a ray which, before entering the crystal, shows no sign of polarisation, has had its varying vibrations analysed, in its passage through the molecular complex forming the crystal, into two plane waves.

When a ray of white light falls on a pile of glass plates with parallel sides, the reflected ray is partially polarised in the plane of incidence, while the refracted ray is partially polarised in a plane perpendicular to this, as must obviously be the case if there is no absorption, and consequently no loss of light in the process of reflection and refraction.

The effect of the transition layer is found to give rise to elliptic polarisation in the reflected ray, the principal axes of the ellipse being parallel and perpendicular, respectively, to the plane of incidence. Some elliptic polarisation is observed, whatever the angle of incidence, and even when the incident ray is plane polarised. The intensity of this elliptic polarisation is greater, the thinner the transition layer.

When a spectrum is obtained by means of a diffraction grating, it is quite obvious that periodicity is introduced by means of the grating, for the radiation constituting any selected portion of the spectrum is composed of contributions from each line or physical element of the grating, and the primary spectrum is the result of an average taken over as many wave-lengths in the incident radiation as there are lines in the grating. If, therefore, two portions which are less than n wave-lengths apart are compared, n being the number of lines in the grating, they will have constituents in common; if less than half n wave-lengths apart, they will have more than half their constituents in common. When dispersion is produced by means of a prism, it is not so easy to form a definite picture of the manner in which the periodicity is introduced, though we can see generally that the effect of the sympathetic molecular vibrations which are the cause of the dispersion will be to collect the contributions making up any one wave-length from a long series of preceding and following wave-lengths, that is to say, that it physically manufactures the periodicity on the same lines as those on which it is manufactured theoretically by the Fourier analysis.

Faraday's discovery of the rotation of the plane of polarisation by the action of a magnetic field is susceptible of very simple explanation; for, consider the plane polarised ray as being composed of two circularly polarised rays in opposite directions, then the vibration which coincides with the natural orbital motion of the electrons will be retarded, while the other will be accelerated, and therefore the plane of polarisation of the resultant ray will turn, as the ray traverses the medium, in the direction of the quicker rotation. It is evident that the amount of turning will, in the case of a given medium, be proportional to the thickness of the medium traversed.

In the general analytical development of the electron theory, the only possible method of representing magneto-optic phenomena consists in the addition of a term in the energy function which will express a connection between the luminous waves and the magnetic field.

A molecular theory taking into consideration the influence of the separate molecules embedded in the medium is again indicated

by the results of physicochemical experiments pointing to the existence of molecular equivalents, both in the case of magnetic rotation, and of the structural rotation observed in the case of many crystals and solutions. Since the magnetic permeability may, as we have seen, always be taken as being equal to unity in the case of luminous waves, the only relation to undergo rotational modification will be that between the electric force and the resulting electrical induction, or electric displacement.

Now, the dispersion which occurs in the rotation, just as in the case of refraction, and which may be similarly accounted for, was shown experimentally by Biot and Verdet to be inversely proportional to the square of the wave-length, and it can be shown to be a mathematical consequence of this that the electric polarisation must be a linear function of the rates of change of the components of the electric force, as well as of the components themselves. In the case of structural rotation these rates of change will clearly be spatial, but in the case of magnetic rotation they will be time rates.

It is of interest to note that this magnetic rotation of the plane of polarisation could not be accounted for by any *molecular* structure in the medium which is only of a statically asymmetric character. A molar structure of this nature would be sufficient to explain such an action, provided it were on a scale comparable with the wave-length of the radiation and were reversed on reversing the direction of magnetisation. Such a method of accounting for the magnetic rotation is, however, excluded by two quite independent considerations. In the first place, a molar structure of this nature would be possible in solid bodies only, and could not exist in fluids, whereas various fluid media exhibit the phenomena of magnetic rotation. The second consideration is one which can here be indicated only in general terms, as its detailed discussion would involve mathematical analysis which is beyond the scope of the present work.

The energy per unit volume of the distribution of polarisation established by an electric force must be a quadratic function of the components of the electric force, and, to a minute extent, of the components of the time rate of the force, the rotation being due to the latter minute portion. It can be shown to follow from this that the effect on the material medium of an external magnetic field or other vector agency is to modify the induced electric polarisation by the addition of a part at right angles to the vector representing the time rate of the electric force, and to a vector whose components are the differences in pairs of the small coefficients of the rotational terms divided by $4\pi c^2$, and equal to their vector product. Now, if the electric force, the

magnetic field, and the time are simultaneously reversed in sign, the effect on the electric polarisation must also be merely to reverse its sign, and therefore a reversal of the magnetic field cannot affect that portion of the energy which depends on the electric force. If, therefore, the coefficients of this part of the energy function are affected at all by the magnetic force, they can only depend on its square or higher even powers. Now, the rotational effect depending on the first power of this force is extremely minute, depending as it does on a small quantity divided by $4\pi c^2$, where c is the constant of radiation. Any rotational effect depending on the square and higher powers of the imposed magnetic force must therefore be quite negligible. This result is confirmed by the results of experiments by Mascart, which show that the mean value of the velocities of a right-handed and a left-handed circular wave-train is sensibly equal to the radiation velocity proper to the medium in the absence of a magnetic field.

It therefore follows generally that, even in a crystalline medium, any dependence, however it may arise, of the electric polarisation on the time rate of change of the inducing electric force, must consist in the addition of a purely rotational part which is symmetrical about an axis, and can only have its origin in the existence of cyclical momenta in the medium, such as are introduced by the revolving electrons in the molecules.

Lord Kelvin has, moreover, shown that statical asymmetry in *molecular* structure cannot under any circumstances account for the optical rotation, as, owing to the fact that a statically asymmetric system is equivalent, as far as pure vibration is concerned, to its optical image, the linear equations of an ordinary elastic medium cannot include any rotational property. The existence of steady cyclical momenta must, therefore, necessarily be postulated, so long as we are concerned with molecules which are so small compared with optical wave-lengths as to act, not directly, but only by sympathetic vibration.

Bose shows experimentally (*Roy. Soc. Proc.*, 1897) that plane polarised radiation, consisting of short Hertzian waves, undergoes rotation of its plane of polarisation when traversing spirally twisted fibrous substances, but the chiral element in this case is a structure of the same order of dimensions as the wave-length of the radiation.

The fact that certain substances, as, for example, quartz crystals and solutions of sugar, when traversed by plane polarised light, cause a rotation of the plane of polarisation in the absence of a magnetic field, demands the existence of asymmetry in structure. Bose's experiments show that a molar asymmetry, in which the

elements of asymmetry were sufficiently large aggregates of molecules, would account for the phenomenon in solid bodies, provided that these aggregates were arranged so as to form a regular chiral structure throughout the substances. This could not be the case in fluids, whether the element of asymmetry were a molecule or an aggregate of molecules, and therefore optical rotation in fluids requires, in addition to the presence of chirally asymmetric molecules, or aggregates of molecules, the existence of steady cyclical momenta, such as are involved in molecules consisting of, or containing, revolving electrons. The entire disappearance of the rotational power of some crystals on fusion shows that in their case, if the asymmetry is molecular, the elements of asymmetry must be aggregates of molecules, and not single molecules. The observed gradual changes in the rotational power of many newly formed liquids strongly suggests that in their case also the effects are due to molecular aggregates. Drude, in his treatise on optics, has apparently overlooked these considerations, as he seeks to account for the effect in all cases simply by asymmetry in the constitution of the molecules. He also appears to overlook the necessity of postulating the existence of steady cyclical momenta in the substances, although this does not invalidate his mathematical argument, as the procedure which he employs implicitly involves their existence.[1] If structural rotation were entirely due to the chirality of the molecules, and not to the chirality of aggregates, the rotational power of a fused crystal ought in all cases to have a value not less than about a third of the maximum value which it possesses in the solid state, supposing each molecule to have a single chiral axis; while if the molecule were equally chiral about all axes, there should be no diminution of the rotational power through fusion. It can hardly be supposed, on the other hand, that the molecules are entirely without intrinsic chirality, since absolutely non-chiral molecules could hardly have the power of spontaneously forming a chiral structure. If this inference is correct, chiral construction of a molecule must be possible without necessarily involving chirality in its vibratory interaction with a radiation. A chiral molecule must necessarily differ from its optical image, and, according to our present point of view, the

[1] *Mathematical Note.*—Drude's mathematical argument is based on the assumption that the motion of an ion depends, not only on the electric force at the place occupied by it, but also on the electric forces in the immediate neighbourhood; so that the first differentials, with respect to the co-ordinates, of the electric force components, must enter into the equations. This, according to Lord Kelvin's result, implicitly involves the existence of steady cyclical momenta. When these conditions are expressed in the most general form physically possible, the resulting equations account for the observed phenomena.

perversion of a molecule into its optical image would be effected by reversing the directions of the orbits of the electrons with respect to the mean configuration in each molecule, which would account for a change from right-handed to left-handed rotation, or *vice versa.*

The existence of chemically isomeric substances, differing only in the direction in which they rotate the plane of polarisation, such, for example, as the dextrorotatory and lævorotatory forms of grape sugar, and which are indistinguishable in their other physical properties, would become explicable. The smallness of the optical effect in comparison with the obvious differences in crystalline form which characterise such isomeric substances is explained by the absence, according to the present theory, of any direct connection between them.

The effect of motion through the ether, both on structural and magnetic rotation, is fully investigated up to the first order by Larmor, who shows that there is no first order effect on the rotational coefficients. This is in accordance with Mascart's experiments on the optical rotation of quartz which were made in 1872, and also with the general conclusion previously arrived at on the basis of the electron theory, that the structure of molecules, and the form of material bodies, is not altered by motion through the ether up to the first order in terms of the ratio v/c, where v is the velocity of motion through the ether, and c is the constant of radiation.

A very important and interesting discovery was made by Zeeman in 1897, which is known as the Zeeman effect. He observed that the bright lines in the spectrum of sodium were perceptibly widened when placed in a strong magnetic field. This result was communicated to Lorentz before publication. Lorentz had, as mentioned in Chapter I., developed an electron theory mathematically in 1892, which, however, received very little attention at the time. On receiving this communication from Zeeman, he was able to show that the phenomenon could be explained on his electronic theory, and also to predict further phenomena, which were afterwards observed, viz. that the lines would be doubled if viewed along the direction of the lines of magnetic force, and trebled if viewed in a perpendicular direction; also that the double lines would be circularly polarised in opposite directions, and that the triple lines would be plane polarised, the plane of polarisation of the central one being perpendicular to that of the two outer ones. The explanation in the first case, when looking along the lines of force, is that, of the two sets of electrons revolving in opposite directions, one set would be retarded and the other accelerated, giving rise to a difference in

wave-length with a consequent difference in the amount of refraction, and therefore in the position of the line in the spectrum. The direction of polarisation shows that the revolving particles are negatively charged. From the results of these observations Zeeman calculated the ratio of the charge of an electron to its mass. It was this discovery which first brought the electron theory prominently into notice; but the high value obtained for the ratio of the charge to the mass led to the results being received with considerable doubt until the result was confirmed by the investigations of Sir J. J. Thomson and others which were considered in Chapter VII.

When the line of view is perpendicular to the magnetic lines of force, one set of electrons will again be retarded and one set accelerated, but the eye will also be affected by electrons vibrating in a direction parallel to the lines of force, and their vibrations will be unaffected by the magnetic field, since they will not cut through any tubes of force. Three lines will therefore be seen instead of two.

It was pointed out by Voigt in 1899, and experimentally verified by himself and Wiechert, that when a plane polarised electromagnetic wave traverses a magnetic field in a direction perpendicular to the lines of magnetic induction, the vibrations, parallel and perpendicular to the field respectively, must be transmitted with different velocities.

Consider a line split by the magnetic field into a triplet, in accordance with the Zeeman effect; the outer components then affect the velocity of propagation of vibrations perpendicular to the field, and the central component affects the vibrations parallel to the field. Approaching an absorption line from the red, there will be a diminution in the velocities of both components, the effect being greater in the nearer one, viz. in the vibration normal to the field. When the absorption line is approached from the violet end, both components will be accelerated, the one vibrating normally being again the one most affected. Double refraction therefore occurs in the sense that at the red end the vibration parallel to the field is propagated with the greatest velocity, and at the violet end the vibration normal to the field.[1]

We have now to consider the mechanical pressure of radiation which was referred to in Chapter VIII.

[1] This elementary theory must be considered as merely an illustration of the character of the phenomenon. See also Foster and Porter's *Electricity and Magnetism*, p. 605. G. A. Schott has shown (*Phil. Mag.*, vol. xiii., 1907, p. 189) that the motions of single electrons would not give rise to sufficiently intense disturbances to form visible lines. The mathematical reader will find a more complete investigation in Appendix F to Sir Joseph Larmor's *Æther and Matter*.

When a beam of light is traversing the ether, there is neither absorption nor reflection, and therefore no arrest or turning back of the energy transmitted by the beam. In material media, however, there is always absorption, and when the beam crosses a practically abrupt interface between the media there will also be reflection, giving rise to a turning back of some of the energy of the waves, with a resulting mechanical traction on the interface. When absorption takes place, some of the energy is arrested in its onward progress, and remains in the medium in the form of unorganised molecular energy of vibration, and this also gives rise to a mechanical traction, acting at every point of the medium at which there is absorption.

A general investigation of the phenomenon would involve, as previously pointed out, the consideration of the polarisation current, and would not be amenable to treatment by elementary mathematical methods. I shall therefore consider only the two ideally simple cases of a wave normally incident from free ether upon a perfectly black body, and a perfectly reflecting one, respectively. Under these conditions Maxwell's procedure would be valid, and we shall therefore arrive at the same numerical results as those given by him, and which are correct, to a very close approximation, whenever light waves proceeding from free ether are incident upon natural bodies which absorb or reflect most of the incident light, besides illustrating the general nature of the phenomenon.

When radiation is incident upon bodies whose dimensions are comparable with, or small relatively to, the wave-lengths of the radiation, there can neither be absorption nor reflection in the ordinary sense. The smallest particles of matter will, however, scatter some of the radiation incident upon them, and so give rise to a turning backwards of some of its energy, and therefore to mechanical pressure, exactly as in the case of true reflection.

That this will be the case is shown by Tyndall's experiments[1] on the colours of precipitated clouds of small particles, and on the blue colour and polarisation of the light from the sky, and Lord Rayleigh's theoretical investigations suggested by them.[2]

The emission of light, mainly of a blue colour, in directions at right angles to the line of propagation, shows that the particles are small compared with the wave-lengths of light, as otherwise they would not give rise to a maximum disturbance in the blue portion of the spectrum. The blue tint of distant mountains is

[1] *Phil. Mag.*, vol. xxxvii., 1869, p. 384; and *Phil. Trans.*, vol. clx., 1870, p. 333.

[2] Lord Rayleigh's *Scientific Papers*, vol. i. pp. 87, 104, 518; and vol. iv. pp. 305, 397.

another example of the same effect. Lord Rayleigh shows that the smallest particles will give rise to scattering of the incident light, there being no finite lower limit to the size requisite for this effect, and in the last paper referred to he adduces evidence tending to the conclusion that the air molecules of the atmosphere are responsible for nearly a third of this scattering, the remainder being probably due to foreign matter.

It is a well-known fact of observation that the much larger particles of which volcanic dust is mainly composed give rise to reddish effects, showing that the disturbance produced by such particles is mainly in the less refrangible, or red, portion of the spectrum. This effect must not be confounded with the observed increase in the redness of the sun as it approaches the horizon, which is due to the increased thickness of the layer of atmosphere traversed by the rays, and the consequent increased elimination of the blue rays by scattering.

We will first consider the case of a plane wave from a constant source of light falling perpendicularly upon a perfectly black body, that is to say, one which absorbs all the rays which fall upon it, neither reflecting nor transmitting any of them.[1] The coefficient of absorption of such a body must be infinitely small, for otherwise the theory of metallic reflection shows that reflection would take place, even when the refractive index is equal to that of the medium, as it must be. Therefore, in order that no light may be transmitted, the body must be infinitely thick. The nearest attainable approach to a perfectly black body is a small hole in a hollow body, for the rays entering the hole will be repeatedly reflected from the walls of the hollow, and only very few will be again reflected out of the hole. The proportion so reflected will evidently be smaller, the smaller the hole in comparison with the surface of the body. Suppose a plane wave-train, travelling in the positive direction along the axis of x, to fall on a perfectly black body, and consider a cylindrical tube of radiation parallel to this axis and of cross section s. Suppose the energy to enter at the point $x=0$. This energy will be completely absorbed, that is to say, converted into heat, within the body, which is supposed to extend to the distance $x=\infty$. Let E be the vibrational energy per unit of volume in the ether adjacent to the body, and c the velocity of radiation in the ether. Then the amount of energy absorbed in time t will be $\mathrm{E}sct$. Now suppose the body displaced through a distance dx in the direction from which the light is coming. The energy falling upon the body in the time t will then be diminished by $\mathrm{E}sdx$, being the amount of energy contained in a length dx of the tube. The amount of heat energy developed in the body will

[1] See Drude's *Lehrbuch der Optik*, p. 447, or English edition, p. 488.

therefore be decreased by this amount. Now the displacement of the body does not affect the amount of vibrational energy entering the tube in the time t, and the electromagnetic energy contained in the length dx of the tube is also unaffected by the displacement; for the index of refraction, and therefore the dielectric constant, is the same for the body and the medium. The electric and magnetic forces are therefore the same at the surface of the body and in the medium. The loss in heat energy developed, arising from the displacement, must therefore, in accordance with the principle of the conservation of energy, be represented by work done in displacing the body, viz. $P_A s dx$ where P_A represents the pressure per unit of area exerted normally on the black body. Therefore

$$P_A s dx = E s dx,$$

so that

$$P_A = E.$$

That is to say, the radiation pressure exerted by plane waves falling perpendicularly upon a perfectly black body is equal to the vibrational energy per unit volume of the external ether.

We will next consider the radiation pressure on a perfectly reflecting body.[1] Consider a wave-train in which the displacement is

$$\xi = a \cos m(x + ct),$$

where c is the velocity of radiation. The direction of propagation is therefore that of decreasing x. Let this wave-train be directly incident on a perfect reflector travelling towards it with a velocity v. The position of the reflector at the time t will be given by the equation $x = vt$. There will then be a reflected wave-train in which the displacement is

$$\xi' = a' \cos m'(x - ct).$$

Since the disturbance does not travel into the reflector, it must be annulled at its surface, so that when $x = vt$, $\xi + \xi' = 0$ identically. This gives

$$a' = -a \text{ and } m' = m\frac{c+v}{c-v}.$$

The amplitude is therefore unchanged, while the wave-length is altered in the ratio

$$\frac{c-v}{c+v} = 1 - \frac{2v}{c}\text{ , approximately,}$$

provided v is small compared with c, and this is the usual statement of the Doppler effect.

[1] See Sir Joseph Larmor's article "Radiation," Supplement to *Encyclopædia Britannica*.

The energy in the wave-train, being half potential and half kinetic, is obtained by integrating $\rho\left(\frac{d\xi}{dt}\right)^2$ along the train, where ρ is the density of the ether. Now we have

$$\frac{\left(\frac{d\xi}{dt}\right)^2}{\left(\frac{d\xi}{dt}\right)^2}=\frac{m^2a^2c^2\left(\frac{c+v}{c-v}\right)^2\sin^2 m\frac{c+v}{c-v}(x-ct)}{m^2a^2c^2\sin^2 m(x+ct)}.$$

Effecting both summations over unit length of the train, the ratio is $(c+v)^2/(c-v)^2$, that is to say, the energy per unit of length is augmented by reflection in this ratio; but the length of the train is diminished by the reflection in the ratio $(c-v)/(c+v)$, so that the energy transmitted per second is increased in the ratio $(c+v)/(c-v)$. This increase can only arise from work done by the advancing reflector against pressure exerted by the radiation. Consider a tube of unit area in the direction of transmission, and let P_A be the pressure of the radiation on unit area of the reflector. Then the work done per second by the advancing reflector is $P_A v$, and the energy impinging on unit area of the reflector per second is equal to the amount in a tube of unit section, and length $c+v$, viz. to $(c+v)E_1$, where E_1 is the energy in unit length of the tube, due to the incident train. Let E_2 be the energy in unit length of the tube due to the reflected train. Then

$$P_A v=(E_2-E_1)(c+v)=\left(\frac{c+v}{c-v}-1\right)(c+v)E_1=\frac{2vE_1(c+v)}{c-v}.$$

Therefore

$$P_A=\frac{2(c+v)E_1}{c-v}=\frac{2(c+v)(c-v)^2E_2}{(c-v)(c+v)^2}.$$

Let E be the energy in unit length of the unit tube due to both incident and reflected trains, so that $E=E_1+E_2$. Then we have

$$E_1=\frac{(c-v)P_A}{2(c+v)},\quad E_2=\frac{(c+v)P_A}{2(c-v)},$$

$$E=\frac{(c-v)^2+(c+v)^2}{2(c^2-v^2)}P_A=\frac{c^2+v^2}{c^2-v^2}P_A.$$

Therefore

$$P_A=\frac{c^2-v^2}{c^2+v^2}E,$$

so that the radiation pressure per unit area is the fraction $(c^2-v^2)/(c^2+v^2)$ of the total density of energy, in front of the

reflector, belonging to both incident and reflected trains. When v is small compared with c,

$$\frac{c^2 - v^2}{c^2 + v^2} = 1 - 2\frac{v^2}{c^2}, \text{ approximately},$$

and this only differs from unity by a term proportional to the square of the small quantity v/c, so that, to the first order, the pressure on unit area is equal to the energy per unit of volume.

We find therefore that the pressure per unit area of normally incident plane waves upon either a perfectly absorbent or a perfectly reflecting body is equal to the density of the vibrational energy in the ether in its immediate neighbourhood. We may therefore conclude that the law holds for all cases in which the whole of the light is either absorbed or reflected, none being transmitted. If a portion is transmitted, then the pressure will be diminished in the ratio of the amount transmitted to the sum of the amounts reflected and absorbed. Now, the work done across a small surface of a wave-front cannot depend on the question whether the wave is plane or not. Therefore the relation must hold good for any simply periodic electromagnetic disturbance, and if P_A be this pressure per unit of area, we shall have, where E now represents the numerical value of the electric force,

$$P_A = \frac{\mu H^2}{4\pi} = \frac{KE^2}{4\pi}.$$

It will be seen from this formula that the radiation pressure upon a body will be proportional to the projection of the area exposed to the radiation upon the plane normal to the direction of the latter, while its weight is proportional to its volume. If, therefore, the linear dimensions of a body are halved, its weight will be reduced to an eighth, while the pressure of radiation will only be reduced to a quarter. It follows, therefore, that if the body is sufficiently reduced in size, a stage will be reached at which the radiation pressure becomes greater than the weight. Reasoning on these lines from the known value of radiation at the surface of the sun, Arrhenius concludes that an opaque sphere of the same density as water would be more repelled by the sun's radiation pressure than attracted by its gravitation, if its diameter were less than a thousandth of a millimetre, and would therefore be repelled from the sun.

CHAPTER XIII.

THE MECHANISM OF RADIATION.

THE simplest method of arriving at a definite physical representation of the manner in which radiation is emitted from material molecules or from isolated electrons is through the consideration of methods which have been actually employed for exciting electromagnetic waves in the ether.

As far back as the year 1842 it was suggested by Joseph Henry that the phenomena accompanying the Leyden jar discharge afforded indications of its being of an oscillatory character; and von Helmholtz, in his celebrated essay on the Conservation of Energy, which was published in 1847, assumed that this would necessarily be the case, and pointed out that the successive oscillations would become gradually smaller, until the entire energy of the discharge was dissipated by the opposing resistances. Lord Kelvin, in 1853, gave a definite mathematical proof of the fact that the discharge must in general be oscillatory, and this was experimentally demonstrated by Feddersen in 1859. The time of a complete oscillation depends on the electrical resistance, the self-induction, or *inductance*, and the capacity of the circuit. When the resistance of the circuit is negligible in comparison with the inductance, a condition which is fulfilled in all the cases considered in this chapter, the time τ will be given by the formula

$$\tau = 2\pi\sqrt{LS},[1]$$

where L is the inductance of the circuit for infinitely rapid vibrations, and S is its capacity, both being expressed in electromagnetic measure. S will then be sensibly equal to the capacity of the Leyden jar, as the capacity of the rest of the circuit will be negligible in comparison with that of the jar. If s represents the capacity of the jar in electrostatic measure, then $S = s/c^2$, where c is the constant of radiation, so that the formula may be written

$$\tau = 2\pi\frac{\sqrt{Ls}}{c}.$$

[1] See Foster and Porter's *Electricity and Magnetism*, p. 441.

That is to say, the distance the disturbance will travel through the free ether during one complete oscillation, in other words, the wave-length λ of the vibrations in free ether, will be given by the formula

$$\lambda = 2\pi\sqrt{Ls},$$

and it will be sensibly the same in the air.

The signification of these electrical constants may be illustrated by a comparison of the electrical oscillation with a mechanical oscillation of a simple character, such as that of a straight steel wire, or a flat strip, fixed at one end, and having a weight attached to its free extremity. The flexibility of the spring will then be the analogue of the capacity, while the inertia of the loaded spring will be the analogue of the inductance of the circuit, and an increase in either of them will diminish the rate of oscillation, and therefore increase the period τ. It should be noted that the complete oscillation, occupying the time τ, consists of a backward and forward oscillation, or a *swing-swang*, as Professor Clifford has called it.

In the electrical case the capacity of the circuit may be increased by increasing the size of the jar. The inductance is due to the magnetic induction in the medium surrounding the circuit, or, more precisely, from our present point of view, to the variation in the electric displacement which accompanies, and is indicated by, the magnetic induction. It can therefore be increased by increasing the length of the circuit. Now, when the electric displacement, and consequently the resulting electric force, is, as in the present case, rapidly alternating in direction, the induced magnetic field is almost entirely confined to the immediate neighbourhood of the conductor. The area included within the circuit has, therefore, very little effect on the inductance, so that the circuit may be made more compact, without sensibly reducing the inductance due to a given length of wire, by winding it into a coil. If the oscillations were slower, the inductance might be still further increased by filling the space inside the coil with a core of soft iron wires; but, in the case of the extremely rapid oscillations produced under the present circumstances, the iron is protected from magnetisation by the currents, opposed to those in the coil, which are induced by the latter in the outer skin of the iron. The introduction of an iron core would, on this account, not only fail to increase the inductance, but would actually diminish it.

When the spring, in the mechanical analogue, is set in motion, the vibrations rapidly die away, primarily owing to the *damping* action of the internal viscous resistance of the material of the

spring to distortion, which causes a dissipation of the energy of vibration into unorganised molecular vibration, in other words, into heat. The damping effect may be greatly increased by immersing the vibrating spring in a viscous medium; and if the viscosity of the medium is sufficiently great, the motion may become *dead-beat*, that is to say, a simple excursion and return to the position of equilibrium. Another cause of damping is the production of waves in the air, or the more viscous medium, surrounding the spring; and if the spring is so shaped as to increase this effect, the damping will be further augmented.

The electric oscillations are damped in a very similar manner, the resistance of the circuit corresponding to the viscous resistance to distortion; and if it is desired to destroy the oscillatory character of the discharge from a Leyden jar or other form of condenser, it will be necessary, except in the case of very large condensers, such as are employed in submarine telegraphy, to include in the circuit some very bad conductor, such as a wet string or a block of wood. The very great capacity of exceptionally large condensers makes such a device unnecessary, as the time of a single swing-swang is so much lengthened that the damping effect of the ordinary resistance of the circuit, continuing during the whole of that time, is sufficient to dissipate the entire energy of the discharge. When the Leyden jar circuit is so designed as to form an efficient exciter of electric waves in the free ether, a rapid damping of the oscillations will necessarily ensue, as a consequence of the principle of the conservation of energy.

The first experimental demonstration of the possibility of exciting electric waves in the ether in this manner was obtained in the celebrated researches of Hertz, the results of which were published in 1888. Several years before this, Professor G. F. Fitzgerald had suggested that such waves might be obtained by means of the discharge of Leyden jars of suitable design; and about the time that Hertz began his investigations, Sir Oliver Lodge was, in connection with inquiries into the action of lightning conductors, making a series of experiments on the discharge of small condensers, which led him on to the observation of electric waves in wires travelling with such great velocities as to show that they were not waves of conduction transmitted by means of the metal, but were conveyed through the ether within the substance of the wires. As Hertz himself has suggested, Lodge would in all probability have succeeded in detecting the ether waves in air, had he not anticipated him. Hertz was experimenting in 1886 with a pair of silk-covered spirals of copper tape similar to those used by Joseph Henry in his researches on mutual and self-

induction about the year 1838. He found, in the course of these experiments, that the large batteries which had hitherto been employed to obtain sparks between the terminals of one of these spirals might be replaced by a Leyden jar, of small size even, provided—and this was the important point—that the discharge was made to jump across a spark-gap. It was this observation which formed the starting-point of the splendid series of researches which afforded an experimental demonstration of Maxwell's electromagnetic wave theory, and formed the foundation for the various systems of ether-wave telegraphy now in such extensive use.

The *oscillator* employed by Hertz as an exciter of ether waves is simply a Leyden jar so designed as to facilitate the transference of the energy of the electric oscillations occurring during its discharge to the surrounding ether. A comparatively large amount of energy is therefore required to maintain it in action, and this

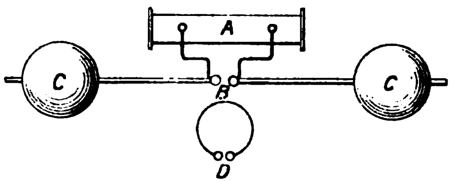


FIG. 21.

was supplied, in Hertz's experiments, from the terminals of the secondary circuit of a powerful induction coil. The arrangement adopted is shown in fig. 21. It consists of a pair of brass rods terminating in small polished knobs B, the distance between which is adjustable, while two large metal spheres or plates CC can be made to slide along the rods. When the coil is set in action, the electric elasticity of the air in the spark-gap B is broken down, so that the air-gap becomes a conductor, and the charges on the spheres C oscillate backwards and forwards from one plate to another, exactly as in the case of the Leyden jar discharge previously considered, except that in the present case there is a continuous supply of energy from the induction coil, and therefore the electrical oscillations can be maintained continuously. These rapid oscillations are only excited when the spark-gap is broken down suddenly. If it breaks down gradually, the oscillations across the spark-gap will consist merely of comparatively slow, almost dead-beat, alternations, corresponding to the alternations in the secondary circuit of the induction coil, and these will be accompanied by an almost continuous glow, or brush

discharge. In order to obtain the required suddenness of breakdown, the interruptor in the primary circuit of the induction coil must be of such a kind as to break and remake the circuit as abruptly as possible, and a large condenser must be employed, in order that the consequent breaking and remaking of the secondary circuit may also be as abrupt as possible. In addition to this, the balls forming the spark-gap terminals should be brightly polished, to avoid a gradual breakdown arising through a brush discharge between the terminals, which is greatly facilitated by their roughness. The balls soon become roughened by the spark discharge, and must then be removed and repolished. The *resonator* D, consisting of a rectangle or circle of wire containing a spark-gap with polished ball terminals, is employed to detect the presence of the electric waves. The two balls are connected by means of a screw, not shown in the diagram, of insulating material, so that the length of the spark-gap is capable of very fine adjustment within certain limits. When the resonator is in the position found to give the maximum sparking across the gap D, the length of this gap is adjusted by means of the screw, and the capacity of the oscillator is altered by sliding the balls towards, or away from, B, until the maximum effect is obtained. The two circuits will then be *tuned* to unison; or, to use Sir Oliver Lodge's term, they will be *syntonic*.

In order to be able to specify conveniently the principal relative positions which the resonator may occupy relatively to the oscillator, we will imagine a straight line AB to be drawn through the centre of the spark-gap B, at right angles to the length CC of the oscillator, and we will call this the central line. There are then three distinct positions which can be occupied by the resonator circuit, without reference to the position in it of the spark-gap D, with the central line passing through the centre of the latter. These three principal positions, as they may be called, are as follows:—

(1) Plane of resonator in plane containing the straight lines AB, CC;

(2) Plane of resonator perpendicular to the straight line AB;

(3) Plane of resonator perpendicular to the straight line CC.

Hertz found in his experiments that—

In position (1) sparks are observed in all positions of D, the maximum effect being obtained when D is nearest to B, and the minimum when D is most distant from B.

In position (2) sparking is a maximum when D is vertically above or vertically below the central line AB, and decreases to zero when D is gradually turned into the plane containing the straight lines AB and CC.

In position (3) no sparks are observed in any position of D.

These experimental results are easily accounted for in accordance with our present point of view, and I shall follow Sir J. J. Thomson in explaining them by the motion of the Faraday tubes by the aid of figs. 22 to 24, the first two of which are copied from those given on pages 394 and 393, respectively, of his *Recent Researches in Electricity and Magnetism*.

In each of the positions of the resonator which I have called the *principal positions*, it will be observed that, except for the resonator spark-gap, the field of the oscillator is symmetrical with regard to the resonator, so that no effects would be produced on the latter if the spark-gap were closed, so as to make the ring continuous. In that case the tubes passing through the centre of the resonator would produce no effect, and in position (1), every tube crossing it above the centre would have its effect exactly counterbalanced by a tube crossing at an equal distance below the centre. The presence of the spark-gap disturbs this condition. To each tube bridging the spark-gap there will correspond a tube touching the circle at the other end of a diameter drawn through the centre of the gap, but the potential difference introduced into the circuit by the latter is indefinitely small in comparison with that introduced at the spark-gap by the former.

In order to form a clear mental picture of what is actually taking place in the ether at each position of a Faraday tube, it is only necessary to bear in mind that a Faraday tube simply represents, by its direction and cross section respectively, the electric displacement at each point of the field, its direction being that of the axis of ether twist, and its cross section measuring the sum of the absolute twists of the ether elements in unit volume.

When sparks are passing freely across the gap B, the Faraday tubes in the vicinity of the oscillator, which were originally perpendicular to the surfaces of the conductors forming the oscillator, and of the wires connecting it with the induction coil, will become deflected towards parallelism with the currents surging backwards and forwards through them, and they will stretch across the spark-gap B in great concentration, and therefore, in the neighbourhood of the central line, the tubes will all be approximately perpendicular to a plane containing this line and which is perpendicular to the straight line CC. This plane may be called the *central plane*. Their mutual repulsion will cause these tubes to move outwards from B, the direction of motion of each tube in the vicinity of the central plane being along a straight line from B lying in this plane and meeting the tube at right angles to its length. The positive directions of the tubes will, with the backward and forward electric surgings,

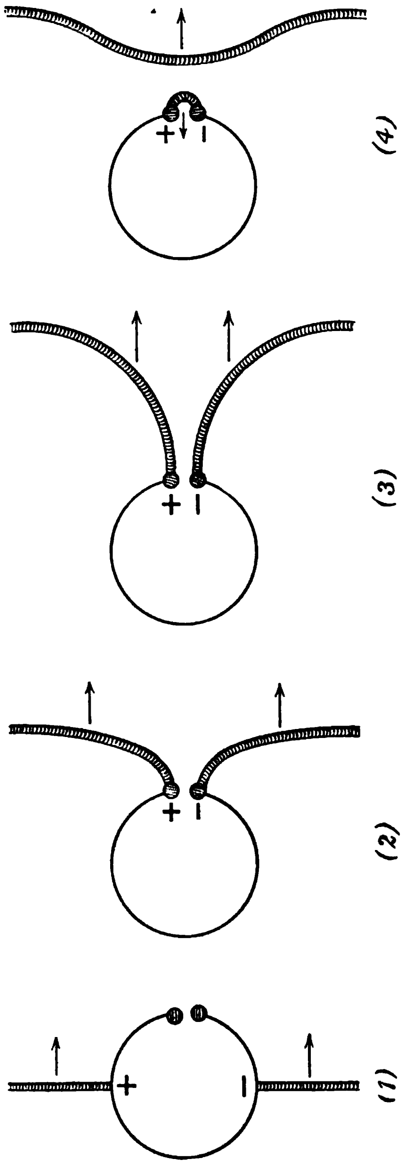


Fig. 22.

alternate from right to left and from left to right, parallel to the line CC.

The portion of the straight line CC included in the spark-gap B, and extending for some distance on either side of it, becomes, in fact, an axis of intense ether twist, alternating with extreme rapidity between right-handed and left-handed twist. This alternating field of twist spreads outwards into the surrounding ether in the form of Faraday tubes or axes of ether twist.

When the resonator is in position (1), with the gap D at the point furthest removed from B, which is the position illustrated in fig. 21, a tube proceeding outward from B, having its length perpendicular to the central line, will strike the resonator circuit at the point opposite to the gap D and divide into two portions. The

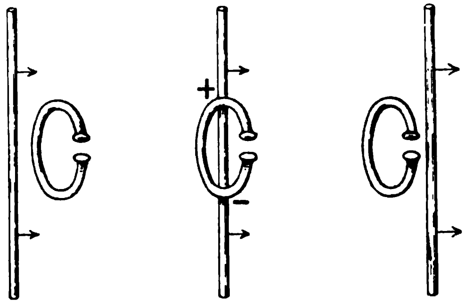


FIG. 23.

end of one portion will travel along one side of the resonator and the end of the other portion along its other side, as shown in fig. 22 (1). As these ends approach the gap, they will bend together across it, as shown successively in (2) and (3). They will then break again, forming a short tube running into the gap and a long tube travelling forwards away from the resonator, as shown in (4).

If the resonator spark-gap were at the point nearest to B, instead of the point farthest removed from it, the gap would be bridged across by a straight tube instead of by the longer curved one of (4), showing that in the former case the electric field at the gap would be more concentrated, and therefore the sparking would be greater.

When the resonator is in position (2), with the spark-gap D vertically above or vertically below the central line, a tube striking the resonator without falling on the spark-gap will split up while crossing the resonator wire, as shown in fig. 23, and will then join

together again and pass on unchanged. Some of the tubes will, however, fall upon the spark-gap, and these will stretch directly across it, and so give rise to a spark. If, while the resonator is still in position (2), it is turned in its own plane through a quarter of a turn, so that the spark-gap is in the plane containing the straight lines AB and CC, the length of the gap will be at right angles to the tubes, as shown in fig. 24, so that none of the tubes will bridge across the gap, and there will be no sparking.

When the resonator is in position (3) the moving tubes, being always perpendicular to the plane of the resonator, can never bridge across the spark-gap, and there will therefore be no position in which sparking can occur.

In every case the tubes passing across the resonator will be

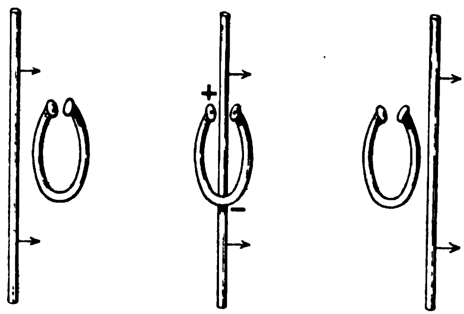


FIG. 24.

divided into groups such that all the tubes of any one group have the positive ends pointing in the same direction, which is opposite to that of the positive ends in the immediately preceding, and the immediately succeeding, group. The maximum effect will clearly be produced on the resonator circuit when these changes of sign in the groups of tubes occur at intervals equal to half the natural period of the resonator circuit, as the changes will then coincide with the reversals of the backward and forward surging in it, so that the effect of the successive groups falling upon the spark-gap will be cumulative. When this condition is not fulfilled, some of the tubes will counteract the effects of others, so that the maximum amount of sparking will not be obtained. Now, the changes of sign in the tubes emitted from the oscillator spark-gap take place at intervals equal to half the natural period of the oscillator, so that the fulfilment of this condition means that the resonator is in syntony with the oscillator.

By calculations made from the readings of an electrometer

inserted in the resonator spark-gap, Bjerkness obtained, in 1891, some interesting curves showing the character of the oscillations. Those arising from a momentary discharge across the spark-gap of an ordinary dumb-bell oscillator (fig. 21) are shown in fig. 25. It will be seen that the damping is extremely rapid, showing the effectiveness of this form of Leyden jar for communicating the electric oscillations to the ether.

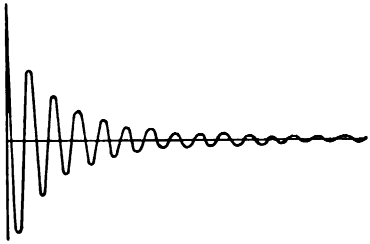


FIG. 25.

The steady character of the oscillations excited in a resonator by an oscillator tuned to syntony with it is shown in fig. 26.

Just as in the case of acoustic resonance, when the resonator has its natural oscillations strongly damped, the tuning of the oscillator into syntony is comparatively unimportant, but if its oscillations are persistent, then exact tuning is essential. When

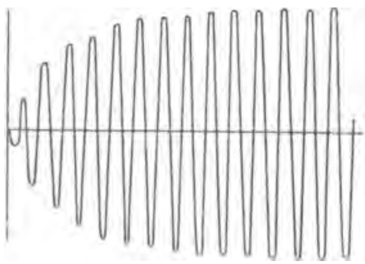


FIG. 26.

the exciting circuit is a persistent oscillator, that is to say, one in which there is very little damping, such, for instance, as an ordinary Leyden jar with its inner and outer coatings connected through a wire, exact syntony is of the greatest importance, as otherwise the exciting circuit will tend to destroy at one moment the oscillations which it set up a moment earlier. This is very well illustrated by fig. 27, which represents the oscillations of a

resonator excited by a persistent oscillation which is not quite syntonic.

Sarasin and de la Rive found, in 1891, that the wave-lengths of the vibrations are very little affected by the dimensions of the dumb-bell oscillator, and, within very wide limits as to these dimensions, are simply determined by the dimensions of the radiator. This is explained by the rapid damping of the vibrations in this form of oscillator as compared with their persistence in the case of the resonator. The difference between the two circuits might have been predicted from a consideration of the distribution of the Faraday tubes just before sparking begins in each case. In the case of the dumb-bell oscillator, many of the tubes stretch across the ether between the two balls or plates CC, and when the insulation in the air-gap B breaks down, many of these will bend round so as to form closed tubes, which will move away from the oscillator with the speed of radiation, carrying

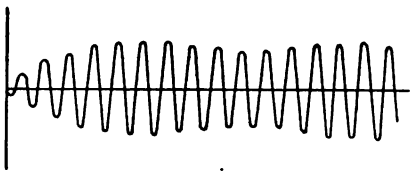


FIG. 27.

away the energy of the oscillator with them. In the resonator, on the other hand, the tubes will stretch from one side of the ring to the other, and when sparking occurs at D, they will merely oscillate backwards and forwards, without any tendency to form closed tubes. There will therefore be little or no radiation of energy, the comparatively slow damping arising mainly from the electrical resistance of the wire.

Hertz employed the resonator described above as a sort of electric eye to determine the presence or absence of electromagnetic disturbance at any desired point in the field of the oscillator. By means of a large flat sheet of zinc at a distance of about a dozen metres from a vibrator with its central line perpendicular to the zinc sheet, the electromagnetic waves emitted by the oscillator, and impinging perpendicularly upon the zinc plate, were reflected back to their starting-point. If this occurred with luminous waves, the direct and reflected rays would interfere with each other, giving rise to a stationary wave with alternations of maximum and minimum disturbance separated by intervals of a quarter of a wave-length. A corresponding periodicity was observed in the points of maximum and minimum sparking, as

shown by exploration with the resonator, along the central line between the zinc sheet and the oscillator.

Hertz also showed that the electromagnetic waves could be brought to a focal line by means of parabolic mirrors, formed by bending sheets of zinc into the form of parabolic cylinders, and that they could be refracted through large prisms and lenses made of pitch.

When the plane waves emitted by the oscillator were allowed to fall upon a grating of parallel wires wound upon a wooden framework, they were found to be entirely transmitted through the grating when the wires were parallel to the Faraday tubes, and entirely reflected when the wires were at right angles to the tubes. The action of the grating on the electromagnetic waves is therefore of the same character as that of a plate of tourmaline on plane polarised light waves.

Light polarised in a plane at right angles to its plane of incidence on the interface between two transparent media, and meeting the interface at an angle whose tangent is equal to the index of refraction from the first medium into the second, is entirely refracted, there being no reflected ray. Trouton, in 1889, succeeded in obtaining a similar result with the waves excited by an electric oscillator. The waves were allowed to impinge upon a thick wall, and reflected waves were observed when the Faraday tubes were perpendicular to the plane of incidence, but no reflection could be detected when the Faraday tubes were in the plane of incidence. This result shows that, in the case of polarised electromagnetic waves, the Faraday tubes are at right angles to the plane of polarisation, which is in accordance with the method adopted in Chapter I. of accounting for the transmission of plane polarised light waves.

The waves employed by Hertz were a few metres long, and later experimenters have succeeded in obtaining electric waves of only a few millimetres in length, but no method has yet been discovered of exciting waves of lengths short enough to be comparable with those of light. If this could be practically effected, it would solve the problem of producing light without heat. The electric arc emits a larger proportion of luminous waves, as compared with heat waves, than any other available illuminant, but even in the electric arc more than 95 per cent. of the total energy is heat energy. The glowworm, the firefly, and many forms of submarine life, have arrived at what appears to be a solution of the problem, but, exclusive of living forms of matter, the nearest approach to it appears to be the faint luminosity which accompanies the slow oxidisation of some substances, of which phosphorus is the most noteworthy instance.

An electrical doublet, consisting of a Faraday tube of molecular dimensions, will constitute an electrical oscillator of much the same type, except for its being on a molecular scale, as the dumb-bell oscillator illustrated in fig. 21. Before, however, considering the nature of the radiation emitted by such a doublet, it will be instructive to consider the simpler method of exciting ether waves, consisting in the sudden stoppage of a moving electron. We have seen (p. 185) that the mass m of the electron is of the order represented by the equation

$$m = \tfrac{8}{3}\pi\rho a^3,$$

where ρ is the density of the ether, and a is the radius of the electron.

Consider now what will happen when the moving spherical

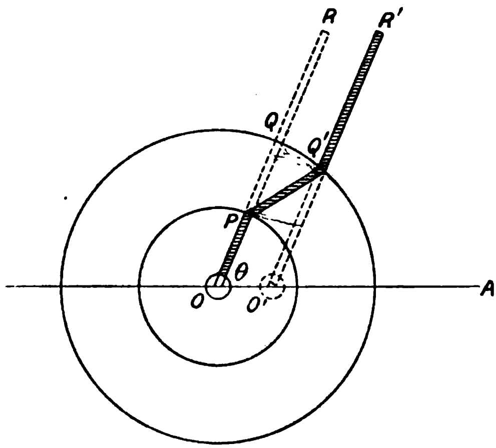


FIG. 28.

electron is suddenly brought to rest, say in a very small time τ. The ends of the Faraday tubes which abut on the electron will first be brought to rest, and the disturbance of the tubes due to the stoppage will travel outwards along them with the speed of radiation. We have to find the configuration of the system of tubes at a time t after the electron has been brought to rest at the point O, fig. 28. From O as centre describe a sphere with radius ct where c is the speed of radiation, and a second sphere with radius $c(t-\tau)$. Then the portions of the tubes outside the outer sphere will occupy the same positions as if the particle had not been stopped, while the portions of the tubes inside the inner

sphere will have come to rest in their final positions. Let O′ be the point which the centre of the spherical electron would have arrived at if it had not been stopped. Let OPQR be the position occupied by a tube at the moment of stoppage. The portion Q′R′ outside the outer sphere will be in the same position at the time t as it would have occupied if there had been no stoppage, that is to say, it will lie on a straight line through O′ parallel to its original position. In order to preserve its continuity the tube must bend round in the thin shell between the two spheres, and will therefore be distorted into the shape OPQ′R′. We will suppose the time τ occupied by the stoppage to be so short that the thickness $\delta = c\tau$ of the shell is very small compared with the radii of the spheres. The portion PQ′ of the tube within this shell may then be considered as straight. We see that within the shell the electric force which was originally radial has now a tangential component E_t, say, and a radial component E_r. Since δ is very small, we may consider $OP = OQ = r$ say, and we shall have

$$\frac{E_t}{E_r} = \frac{QQ'}{PQ} = \frac{OO' \sin\theta}{\delta} = \frac{vt \sin\theta}{\delta},$$

where v is the velocity of the electron at the moment of stoppage, and θ is the angle between OP and the direction of motion OA. But

$$E_r = \frac{e}{r^2} \text{ and } r = ct.$$

Therefore

$$E_t = \frac{ev}{c} \cdot \frac{\sin\theta}{r\delta}.$$

These tangential Faraday tubes moving outwards with a velocity c will produce at P a magnetic force at right angles to the plane of the diagram and in the opposite direction to the magnetic force existing at P before the stoppage of the particle, and of magnitude

$$H = cE_t = \frac{ev \sin\theta}{r\delta},$$

which exceeds the magnetic force $\dfrac{ev\sin\theta}{r^2}$ previously existing at P, in the ratio of r to δ. The outward travelling shell of thickness δ is therefore the seat of intense electric and magnetic forces which only diminish inversely as the distance from the charged particle, whereas those existing before the stoppage diminished inversely as the square of the distance. It is by the sudden stoppage of cathode ray particles in this manner that Sir J. J.

Thomson, to whom this investigation is due,[1] explains the formation of the Röntgen rays, as originally suggested by Sir George Stokes. Rays consisting of single thin pulses of this kind, in the place of trains of radiation, would not undergo refraction, and the great penetrating power of the Röntgen rays is also satisfactorily accounted for. A single thin pulse of this kind will simply impart a sudden impulse to each electron which it passes over in traversing a material body, acting on it very much as if it were entirely isolated from its neighbours. The interactions of the electron, maintained through the ether, with other electrons in the same molecule, being, as Larmor observes, elastic forces depending on the strain, will remain finite, and so will not be in a position to ease off materially the velocity communicated by such impulsive action. The loss of energy by communication from the pulse to the electrons will therefore be simply proportional to the number of electrons per unit of volume, irrespective of their molecular arrangement. This is in accordance with observations made by Röntgen and others that the order of opacity of bodies for Röntgen rays is generally about the same as the order of their densities. He also points out that a single one-sided impulse of this kind would communicate more energy to an electron lying in its path than a forward impulse immediately followed by an equal backward one, so that the Röntgen rays arising from sudden stoppage of the cathode ray particles should be less penetrating than those arising from shock to the electrons belonging to the molecules of the walls of the tube.

In either case the effect is much smaller than that of a wave-train arising from a regularly vibrating molecule, since such a wave-train gradually establishes by resonance, in the course of a large number of periods, a state of molecular vibration involving all the electrons in the molecule affected, in a connected manner, as members of a single dynamical system. If the particle were moving at the moment of stoppage with a velocity nearly equal to that of light, the tubes before stoppage would all be concentrated almost into the equatorial plane. Assuming the limiting case to be possible, Sir J. J. Thomson points out that the configuration of the tubes after a time t may again be determined by finding the configuration which they would have occupied if they had not been stopped. The tubes would in that case have been in a plane BC (fig. 29) at a distance ct in front of the particle. Two spheres are again drawn with radii ct and $c(t-\tau)$ respectively; outside the larger sphere the configuration will again be the same as if there had been no stoppage, that is to say, the tubes will lie in the plane at a distance ct in front of the particle, and therefore touching the

[1] *Electricity and Matter*, by Sir J. J. Thomson, chapter iii.

outer sphere. Within the inner sphere the Faraday tubes will be uniformly distributed, so that, in order to preserve continuity, the tubes must run round within the shell as shown. There will therefore in this case be two pulses, one a plane pulse travelling in the direction in which the particle was originally moving, and the other a spherical pulse travelling outwards in all directions.

Now suppose, with Larmor, that an electron e starts from a point A, and travels over a small distance with the velocity v, arriving after an interval of time δt at a point B, then $AB = v\delta t$, and the effect of its change of position will be the same as that of the creation of an electric doublet AB of moment $ev\delta t$.

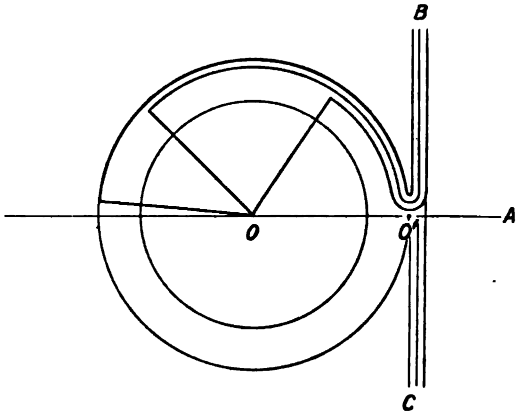


FIG. 29.

Now, except at places whose distance from the nucleus of the electron is so small as to be comparable with the linear dimensions of the nucleus, the electron may be treated as a point carrying a charge. The ethereal disturbance arising from the motion of an electron can therefore be obtained by mere addition of the elementary disturbances due to its passage along the successive elements of its path. It is merely necessary, therefore, to find the disturbance due to the creation of such a doublet, and then to effect a summation of the result along the paths of all the electrons of the vibrating atom. The sudden creation of the doublet of moment $ev\delta t$ is found to give rise, as in the case of the stoppage of a particle investigated above by elementary methods, to a magnetic force varying inversely as the distance; and in addition to this, a magnetic force varying inversely as the square of the distance, forming the permanent magnetic field of the

doublet, and corresponding to the previously existing magnetic field of the uniformly moving charged particle.

The solution is obtained in the form

$$H = -\sin\theta\delta t\left(\frac{ev}{r^2} + \frac{e\dot{v}}{cr}\right),$$

where, as before, H is the magnetic force at a distance r from the doublet, inclined to AB, the direction of motion, at an angle θ. The quantity $\dot{v}$ is the acceleration of the electron whose velocity is v, so that $\dot{v}$ stands for the time rate of v, a notation originally employed by Newton.

For points near to the doublet, the first term will be the more important, the second being comparatively small, owing to its containing the radiation constant in the numerator. As v increases, however, the first term decreases more rapidly than the second, so that at a considerable distance the second term is the important one. For points near to the origin of the disturbance, the magnetic force at a distance r will be in the plane perpendicular to r, at right angles to the projection v', on that plane, of the velocity of the electron, and equal in amount to ev'/r^2. Adopting the interpretation of magnetic force as consisting of an actual ether flow, which corresponds to the assumption that the vibrations are transmitted by actual displacements of the particles of ether, it will follow that the actual path of a vibrating ether particle, at a point near the origin, will be in the plane perpendicular to the radius vector joining the point to the origin, and will be similar to the projection of the electron's orbit on that plane, when turned through a right angle, the magnetic force being, as we have seen, always at right angles to the direction of motion of the electron. This is on the assumption that the wave-length is large in comparison with the size of the orbit of the electron within an atom; as if this were not the case, there would be a lag of phase, the result of which would be that a simple harmonic vibration of the electron, such as uniform motion in a circle, would give rise to a more complex form of vibration. Each line of the spectrum would, in that case, be accompanied by a system of lines due to the various harmonic vibrations into which it could be resolved.

If the acceleration of an electron ceases, so that it moves with uniform speed in a straight line, no more radiant energy will be emitted from it. The sheets of radiation previously emitted will, however, continue to travel on through the ether, leaving behind them, ready formed, the steady magnetic field of the moving electron, which will be the same as would arise from a steady current element ev. The field which becomes established in this manner, as a trail of the shell of radiation due to the starting of

the electron into motion, does not involve any sensible amount of energy, except in the immediate neighbourhood of the electron. When a moving electron is stopped, its steady magnetic field will be wiped out by the shell of radiation travelling outwards as a consequence of the stoppage.

At distant points in the ether the steady magnetic field corresponding to the first term in the expression for H will be insensible, so that the effect will be sensibly that due to the second term. If this term vanishes, there will therefore be no appreciable radiation, and this result of Larmor's analysis shows that an atom containing a number of revolving elections will not suffer dissipation of its energy by radiation due to acceleration or retardation of its motion, provided that the sum of the second terms, corresponding to all the electrons in the atom, has the value zero. As these are vector quantities, it is the vector sum $\Sigma e\dot{v}$ that must vanish. This is therefore the necessary and sufficient condition for the permanent existence of an atom. A simple electric doublet, consisting of equal positive and negative electrons revolving round each other, would not be a possible type of atom, for the accelerations of the two constituents would reinforce, instead of cancelling, each other's effects.

Now, this second term is obtained from the first by multiplying by r/c, which is a mere numerical quantity, and by replacing the vector quantity v, representing the velocity of the electron, by the vector quantity $\dot{v}$, representing its acceleration. The projection of the electron's orbit on the plane perpendicular to the radius vector r, that is to say, on the plane of the wave front, will therefore no longer be a curve similar to the path of a vibrating ether particle. The latter path will, however, be similar to the projection on this plane of the path of a particle moving so that its velocity at every moment is numerically equal to, and in the same direction as, the acceleration of the electron, at the corresponding moment, in its orbit. This curve is called the *hodograph* of the orbit. It can easily be shown that the hodograph of a uniformly described circular orbit is also a uniformly described circle, and that the hodograph of a harmonically described ellipse is another harmonically described ellipse. It follows therefore that the ether disturbance due to simple harmonic motion of an electron, whether at near or distant points, is always simple harmonic. That is to say, the intrinsic periods of the radiation are those of the system of electrons to which it is due.

When, instead of considering merely the results arising from the creation or extinction of a doublet, the aggregate contributions to the magnetic force, arising from an electron executing simple harmonic motion, is analytically examined in further detail,

Larmor shows that the first term in the expression for the magnetic force gives rise to no permanent loss of energy from the system, although there may be oscillations of kinetic energy backward and forward through any finite range, and the same holds good for the potential energy, which depends on the electric force. The absence of any resultant loss of energy from backward and forward pulses of magnetic and electric force is obviously conditional on the absence of absorption in the medium, which is true in the case of the ether, but not strictly so for even the most transparent of material substances.

This consideration is of great importance in respect to its effect upon the penetrability of material bodies by electric waves of various lengths. When radiation impinges on an interface between two media, it usually gives rise to at least two trains of radiation, one forming the reflected ray, and the other forming the refracted ray. It might be imagined that, except in the case of waves of periodicities corresponding with some of the intrinsic periods of the molecules of the medium, the proportion of refracted radiation absorbed in penetrating to a given depth would depend simply on the number of vibrations in reaching this depth. The penetrating power would then be simply proportional to the wave-length, so that an electric wave a couple of thousand miles in length would pass through the earth as easily as a light wave would penetrate to a depth of a few thousandths of a millimetre, and electric waves a few metres long, or less, would travel through water to any distance attainable in the ocean. The long waves would, however, continue to emit the backward and forward pulses referred to above for as many wave-lengths as the short ones, and every pulse would be subject to absorption during its passage in each direction.

The second term in the expression for H gives rise to a uniform stream of radiation in the direction of propagation, and its amount is such that the radiation of an atom, for which $\Sigma e\dot{v}=0$, is less than that of one of its electrons alone, moving with the same speed, in the order of the square of the ratio of the diameter of the molecule to the wave-length of the radiation. Since the radiation emitted by material atoms has wave-lengths at least a thousand times the diameter of the atoms, the radiation actually emitted by an atom must be less than a millionth of the aggregate of the radiations that would be emitted by its separate electrons, and enormously less, according to the accepted view of the smallness of the electrons in comparison with the distances by which they are separated within the atom. The ratio of the orbital velocities of the electrons to the speed of radiation must be of at least the same order of smallness as that

of the atomic dimensions to the wave-length, and this may be accounted for, as Larmor points out, by the consideration that the energy of groups of electrons with greater orbital speeds would be sensibly dissipated by radiation, and such atoms would not, therefore, be capable of permanent existence.

We saw in the preceding chapter that the molecular oscillations are not truly simple harmonic, nor even accurately periodic, the simple harmonic vibrations corresponding to the various lines of the spectra being the result of a physical analysis of the wave-trains by the prisms or grating employed to obtain the spectra. The vast number of lines in the spectra of some molecules does not, therefore, indicate any necessarily corresponding complexity in the constitution of the molecules. Larmor very aptly illustrates this by the case of the astronomical system of the Sun, Earth, and Moon, which is determined by nine co-ordinates, three for each body, and yet there exists the much larger number of periodic inequalities, or oscillations, which are dealt with in the Lunar Theory.

A difficulty which requires consideration arises from the conclusion, which we have seen to be essentially involved in the electron theory, that the radiation emitted by material atoms, and by the groups of atoms constituting molecules, is almost entirely due to disturbances in their steady motions, and not to any sensible extent to the steady motions themselves.

It was pointed out by Lord Rayleigh in 1906 that in this case simple relations should exist between the squares of the frequencies corresponding to the various spectral lines arising from the vibrations of similar molecules, and not between the frequencies themselves. Now, it has been shown by numerous series of observations carried out by Rydberg, and also by Kayser and Runge, with great refinements of detail, that simple relations exist between the frequencies themselves, and not between their squares, constant differences of frequency being found to exist throughout numerous series of lines.

Lodge has suggested that this difficulty may be overcome by supposing the perturbation due to chemical or other atomic collision to be of the nature of the disturbance of the centre of gravity of a revolving system. We have seen that the absence of radiation in the case of the motion of an undisturbed atom depends on the null value of the vector sum $\Sigma e\dot{v}$ for all the electrons in the atom. Now, the result of a collision, if it did not permanently alter the constitution of the atom, would be to displace these electrons from their relative arrangement in which $\Sigma e\dot{v}=0$, and would, at the same time, cause a rotation of the centre of gravity around its original position. This rotation would soon be damped

out by the radiation, and the atom would then return to its original condition. During the period within which the radiation continued, the vibrations of the atom might be regarded as arising from the steady orbital motions of the electrons, each perturbed by the motion of the centre of gravity of the system; and if the atom were not broken up, the effects due to the perturbations would in all probability be restricted to comparatively small values in comparison with those due to the steady motions. In this case the train of radiation would consist sensibly of the various harmonics due to the steady intrinsic periods of the atom, as Lodge suggests.

Sir J. J. Thomson's investigations, which are considered in Chapter XXI., afford support to such an explanation as a partial account of what occurs during a collision, and the matter will be referred to again at the end of that chapter. The relation between intensity of radiation and temperature is expressed by Stefan's law, which was originally formulated empirically as the result of experimental observation. According to Stefan's law, the total energy of radiation from a hot body is proportional to the fourth power of the absolute temperature T of the body, so that it may be expressed in the form σT^4, where σ is a constant. Boltzmann has deduced Stefan's law by thermodynamic reasoning from the fact demonstrated at the end of Chapter XII. that the pressure of normal radiation on an opaque body in free ether is equal to the density of the radiant energy just in front of it. The reasoning, however, depends on the assumption that the second law of thermodynamics, or Carnot's principle, can be extended to the ether. J. H. Jeans shows, however, that the second law of thermodynamics is not sufficiently general to be applicable even to different kinds of ideal matter, so that it certainly cannot be considered legitimate to apply it to the ether on the basis of its being experimentally known to be applicable to ordinary matter. Assuming the radiation to be dependent, among other circumstances, on the charge e of each electron of which the matter consists, he shows (see Appendix M) that σ is proportional to e^{-6}, and that the numerical value of e, deduced from the observed value of σ, supplies strong confirmation that this is the right method of evaluating σ.[1] Now suppose that R is a piece of ordinary matter, the electrons of which each have a charge e, and that I is a piece of ideal matter, each electron in which has a charge e' where e' is less than e. If R and I, supposed to be initially at the same temperature, are placed together in a perfectly reflecting enclosure, then the ideal matter will emit more energy than it receives, and the real matter will receive more

[1] See also Chapter XIV., p. 320.

energy than it emits, so that the ideal matter will fall in temperature, while the real matter rises. Again, suppose we take two bars R and I of equal cross section, and suppose them bent and joined into a ring, the contacts being at A and B (fig. 30). Let each face of the junction A have perfect conductivity and no radiating power, while each face of the junction B has no conductivity and perfect radiating power, and suppose the remainder of the surfaces of R and I to be impervious to energy. If the junction A only is thrown out of action, R and I will assume different temperatures.

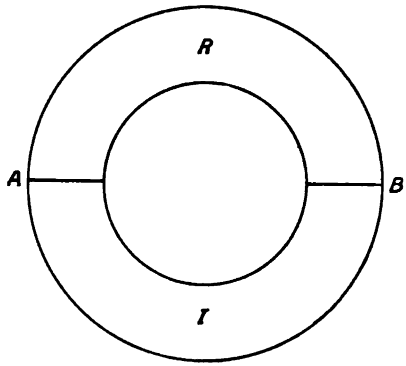


FIG. 30.

If the junction B only is thrown out of action, any difference of temperature will be equalised by conduction across the junction A. If both junctions are in operation together, there will be a continuous flow of energy round the ring in the direction IARB, and a continuous rise of temperature as we pass through I from B to A, or through R from A to B. The latter illustration was suggested to Jeans in conversation with Lorentz. The inference from both is that the second law of thermodynamics is not applicable to different kinds of matter, and still more is it illegitimate to argue, from its applicability to ordinary matter, that it may therefore be extended to the ether.

CHAPTER XIV.

METALLIC CONDUCTION AND RADIATION.

It was shown in Chapter XI. that the unequal mobility of the positive and negative ions does not prevent the two classes of ions having equal shares in carrying a steady current in an electrolyte, but that it necessarily involves a compensating drift of the mass of electrolyte in the direction of motion of the more mobile ions. If the circuit is partly metallic a similar accumulation must occur in this portion unless one of two conditions be fulfilled. Either the two kinds of ions or electrons must have equal mobilities, or the mobility of those which act as carriers of the current must be so great as to make the accumulation insensible.

All the theories that have been put forward to explain the process of metallic conduction are founded, as far as the author is aware, on the assumption that the transmission of electricity through the metal is effected by the movement of electrically charged particles. A theory of this kind was developed by Riecke, and subsequently elaborated by Drude, in which it was assumed that both positively and negatively charged particles were present in metals, and were capable of carrying the electric current by moving through the mass of the metal under the action of electric force. This theory was modified some years ago by Sir J. J. Thomson by limiting the movement to the negative electrons, and assuming that the heavier positive particles took little or no part in the process of conduction. This modification of the theory was adopted in order to bring it into accordance with the fact that no transport of the metal has ever been observed as an accompaniment to electric conduction, although this has often been looked for. If the current were carried to any considerable extent by means of the positively charged atoms, it would appear as if such a result must be observed; for example, if a current were made to flow through a circuit composed of two or more different metals, an alloy should be formed at the point of junction, in one or other of the

metals, according to the direction of the current. A composite conductor of this kind is, as a matter of fact, employed in thermoelectric batteries. These are formed of a number of alternate strips of antimony and bismuth, so arranged that every alternate junction can be heated by means of a gas flame, or otherwise, while the other junctions are kept cool. Such batteries are always found to deteriorate after a certain time, evidently owing to some molecular change going on in the interior of the metallic circuit, and it has been suggested that this change may possibly consist in the production of traces of alloy at the junctions by the migration of positively charged atoms carrying a small proportion of the current. If, however, this occurs, the proportion so carried must be extremely small, as the most careful chemical examination has failed to show any evidence of such a production of alloy, and it is more generally considered that the deterioration is due to the frequent alternations of expansion and contraction. Users of these batteries, moreover, seem in general to be of opinion that this deterioration is less marked when the battery is kept continuously in operation, which tends to support the latter explanation rather than the former one.

In the original form of the theory the electrons were supposed to be split off from the atoms by the action of one atom of the metal on another, and to remain diffused throughout the mass of metal, which might therefore be regarded as having somewhat the character of a porous body with its pores occupied by a substance having the properties of a perfect gas. It was also assumed that the time during which the electrons remained in this condition of freedom bore a sufficiently high ratio to the time during which they were bound within the atoms to enable them to behave like a gas under such circumstances by attaining the same temperature as the metal. The average kinetic energy of an electron would in that case depend only upon the temperature.

Instead of making these assumptions, I shall follow J. H. Jeans,[1] and simply assume that, in a conductor which is under the influence of no externally impressed electric force, there are N electrons in unit of volume moving with velocities distributed in accordance with Maxwell's law,[2] at a temperature determined by their own kinetic energy. Maxwell's law gives the partition of velocities for particles in collision, or acted on by any field of force, as well as when on a free path. The distribution of velocities will be unaffected either by external or intermolecular forces, but these will alter the distribution of density.[3]

[1] "The Motions of Electrons in Solids," *Phil. Mag.*, vol. xvii., June 1909.
[2] See Appendix M.
[3] *Dynamical Theory of Gases*, J. H. Jeans, pp. 70 to 78.

The kinetic theory of gases was first applied to the equilibrium of the free electrons in a metal by Drude,[1] and subsequent experimental researches have shown that the similarity between a free electron in a metal and a gaseous molecule under ordinary conditions is extremely close. By measuring the variation in the resistance of bismuth in a magnetic field, Sir J. J. Thomson has shown[2] that, on the ordinary assumption of collisions and free paths, the mean free path of an electron in that metal is 10^{-4} centimetre, and Patterson[3] finds from experiments on gold, platinum, silver, copper, zinc, tin, cadmium, mercury, and carbon, that the mean free path has values lying between $5{\cdot}9 \times 10^{-7}$ and $4{\cdot}1 \times 10^{-7}$ centimetre. O. W. Richardson[4] points out that since the mean free path for a nitrogen molecule in air at standard pressure and temperature is 10^{-5} centimetre, it follows that in bismuth the mean free path of an electron is about the same as that of a molecule of air at $\frac{1}{10}$th of the standard pressure, while in the other substances experimented on it is about the same as in air at a pressure of ten atmospheres.

It will assist the reader to follow the argument if we take a concrete case; say, for example, the passage of a current of electricity through a copper wire. Approximate determinations of the size as well as the mass of atoms and molecules of different substances have been obtained by Lord Kelvin and others, in many widely differing ways, with results which agree sufficiently well with one another to enable us to arrive at a fair approximation to the number of atoms present in a cubic centimetre, in the case of a large number of substances.[5] The number of atoms of copper in a cubic centimetre of that material would be something of the order of a hundred thousand trillions. From some researches of Schuster's it would appear that from one to three electrons at a time may escape from any one atom of a metal, this number varying with different metals.[6] In the case of copper he estimates that there are about three mobile electrons to every two atoms, and, according to this estimate, a cubic centimetre of copper would contain about a hundred and fifty thousand trillions of free electrons, which would be moving in all directions at an average velocity of about sixty miles a second. With no external electric force acting on the copper, there would be no resultant flow of electricity, as the effect of the number of electrons moving in any

[1] Drude's *Annalen der Physik u. Chemie*, vol. i. p. 572.
[2] *Congrès International de Physique*, Paris, 1900, vol. iii. p. 138.
[3] *Phil. Mag.* (6), 1902, vol. iii. p. 655.
[4] *Phil. Trans.*, A., vol. cci., 1903, p. 499.
[5] See Thomson and Poynting's *Heat*, p. 148, and Jeans' *Dynamical Theory of Gases*, chapter xix.
[6] *Phil. Mag.*, vol. vii., 1904, p. 154.

one direction would be balanced by that of an equal number moving in an opposite direction. The application of an external electric force would, however, at once cause a drift of electrons in a direction opposed to that of the force, owing to their negative electrification, and this drift would constitute an electric current.

We shall see that the theory here set forth will account for the principal phenomena accompanying the flow of electric currents in metallic conductors. It was pointed out in Chapter IX. that in ordinary currents of conduction the flux of electric displacement is negligible in comparison with the electron flux. The electron flow must therefore approximate closely to a stream vector, or, in other words, must sensibly have the characteristics of the flow of an incompressible fluid. There must therefore be no sensible accumulation of electrons at any part of the metallic circuit, a condition which will be fulfilled if their mobility is sufficiently great, as such an accumulation will then be prevented by their mutual repulsions. Now, according to Ohm's law, the strength of a current between any two points of a metallic conductor is proportional to the electromotive force between the two points divided by a certain quantity depending only upon the dimensions and material of the conductor, which is called its electrical resistance. Then, as shown in Appendix M, this law will follow without making any assumptions as to the nature of the motion of the free electrons, and if κ is the conductivity for steady currents, k the conductivity for currents of frequency p, m the mass of a free electron, and e its charge, it is found that

$$k = \frac{\kappa}{1 + \frac{\kappa^2 p^2 m^2}{N^2 e^4}} = \kappa\left(1 - \frac{\kappa^2 p^2 m^2}{N^2 e^4} + \ldots\right).$$

Sir J. J. Thomson[1] obtains a formula which would lead to the expression

$$k = \kappa\left(1 - \frac{1}{3}\frac{\kappa^2 p^2 m^2}{N^2 e^4} + \ldots\right).$$

The discrepancy is attributed by Jeans to Thomson's system being what might be described as a kinematical, rather than a dynamical, system, in that it neglects to take account of the fact that the time of a free path must depend on the velocity with which it is described. The difference is of fundamental importance, since some experiments by Hagen and Rubens on the reflection of long waves from the surface of metals, in conjunction with a formula expressing the variation of conductivity with frequency, gives, as we shall presently see, information as to the

[1] *Corpuscular Theory of Matter*, p. 84.

value of N. The value of N obtained from Thomson's formula would necessitate values for the specific heats of the metals experimented with much greater than their observed amounts, and this led Thomson to reject his earlier theory that the electrons exist in a free state within the metal for a sufficient length of time to attain the temperature of the metal. He replaces this by the assumption that the electric force, instead of acting on the electrons after they have left their atoms, really acts upon the molecules before they have lost the electrons, and arranges them in a sort of Grotthus chain, with their negative ends all pointing in one direction and their positive ends in the opposite one. The result would be that the electrons, instead of wandering about, as when no electric force was acting on the metal, would jump straight from one atom to the next one. This modification would obviate the difficulty of the unduly high specific heat, since the jump of the electron from one atom to the next would occupy so little time that the electrons would not remain free long enough to attain the temperature of the metal, and it would, therefore, never be necessary to consider the amount of energy required to raise the temperature of a large collection of them. The theory is worked out on the assumption that a metallic conductor contains a large number of such doublet molecules capable of moving freely through its substance, and which may therefore be assumed to distribute themselves in accordance with Maxwell's law.[1] It is shown to account for the phenomena of metallic conduction, but these hypotheses are of an extremely artificial character, and the theory based upon them is only propounded by Sir J. J. Thomson as a possible alternative to the earlier one,[2] in consequence of the supposed breakdown of the latter. As Jeans' correction of the formula in question entirely obviates the difficulty which led to the formulation of this alternative theory, I shall not consider the latter in detail, but proceed to show how the phenomena of metallic conduction may be accounted for on the far simpler and more satisfactory theory which assumes only the presence of free negative electrons distributed in accordance with Maxwell's law, at a temperature determined by their own kinetic energy. We shall see presently that this temperature is what is called the temperature of the matter forming the substance of the conductor.

The free electrons will collide with each other and with the atoms of the metallic conductor, and when a steady state is attained, the mean kinetic energy of the electrons, which determines the temperature, will be independent of the pressure, and

[1] *Corpuscular Theory of Matter*, chapter v.
[2] *Ibid.*, chapter iv.

of the masses of the electrons. The mean kinetic energy of an electron, which is measured by half the product of its mass into the square of its velocity, will therefore be the same as the mean kinetic energy of a molecule of hydrogen at the same temperature. A molecule of hydrogen, however, contains two atoms, each having about 1700 times the mass of an electron, so that the mass of the molecule will be 3400 times that of an electron. The mean value of the square of the velocity of an electron will therefore be 3400 times that of a molecule of hydrogen, or the average velocity of an electron will be about fifty-eight times as great as that of a molecule of hydrogen at the same temperature. Now the average velocity of a hydrogen molecule at 0° C. is about $1{\cdot}7 \times 10^5$ centimetres a second, so that that of an electron will be about 10^7 centimetres a second, or about sixty miles a second, as given in the example of the electrons in a copper wire.

When there is no electric force acting on a conductor, there will be no resultant flow of electricity arising from the motion of the negative electrons, but the application of an external force in any direction will cause a resultant drift along its line of action. The change in the motion of the individual electrons by the action of any electric forces at our disposal will be extremely minute relatively to their average speeds; but, owing to the enormous number in every unit of volume, and to the drift being in the same direction for all, a very small drift is sufficient to give rise to a very large current density. Since electric force is taken to be positive when it gives rise to a flow of positive electricity in its own direction, the drift of the free electrons, these being negative, will be in the opposite direction to that of the force. Let us suppose that the electric force X per centimetre is applied parallel to the axis of x in the positive direction, giving rise to a drift $-u$ of the negative electrons in the direction $-x$. Consider a section of unit area at right angles to the axis of x. The excess in the number of electrons crossing this section in the direction $-x$ over the number crossing in the opposite direction x will be Nu per second, giving rise to a flow of electricity Neu per second $=C$, say. It is shown in Appendix M, without making any assumptions as to the nature of the motion of the electrons, that this flow will be in accordance with Ohm's law for any electric forces within our experience. On the assumption that the motion of an electron consists of free paths and collisions, this is easily proved in an elementary manner. In that case the electron will only be sensibly influenced during its free path, the shock of a collision being too violent for its initial velocity and direction immediately following a collision to be sensibly affected by the

electric force. Jeans has shown, however,[1] that the motion immediately after a collision is affected to a certain extent by the previous collisions, the effect of this *persistence* being to increase the mean velocity during the free path in a ratio depending on the structure of the particle and the nature of the action which we call a collision. As in the case of an electron we know nothing about either of these, all that we know about the persistence factor is that it is greater than unity. Calling this factor β, the velocity of the electron, due to the force X, will increase from zero at the beginning of its free path to $\beta X \frac{e}{m} t$ at the end, where t is the interval between two collisions, giving $\frac{\beta}{2} X \frac{e}{m} t$ as the mean velocity $= u$, so that $C = N \frac{\beta}{2} X \frac{e^2}{m} t$. If l is the length of the mean free path, and v the average velocity with which it is described, we may write $l = vt$, neglecting the effect of the electric force in changing the velocity, which gives us

$$C = \frac{N\beta X e^2 l}{2mv} = \frac{N\beta X e^2 l v}{2mv^2}.$$

Now mv^2 is twice the mean kinetic energy of an electron, and therefore equal to twice the mean kinetic energy of a molecule of hydrogen at the same absolute temperature T. But the value of this is known to be 3RT, where $R = 9{\cdot}3 \times 10^{-17}$, being the constant of the kinetic theory of gases, which may be considered as approximately defined by the approximate laws of Boyle and Charles, according to which the pressure of a gas at the absolute temperature T, and containing N molecules per unit of volume, is equal to NRT. Substituting this value for mv^2, we have

$$\frac{C}{X} = \frac{N\beta e^2 l v}{6RT}.$$

But C/X is the specific conductivity of the conductor, and its value is independent of the electric force, which is Ohm's law.

If X could be made large enough to produce a sensible change in the velocity of an electron during the interval between two collisions, Ohm's law would no longer be true, as a term which was not proportional to X would then be introduced into the expression for the current.

Let v' be the velocity so produced, then the kinetic energy arising from the force Xe acting on an electron during the mean free path would be

$$\tfrac{1}{2} m v'^2 = Xel,$$

[1] *Dynamical Theory of Gases*, p. 236.

and therefore the average velocity would be

$$\frac{v'}{2}=\sqrt{\frac{eXl}{2m}}.$$

This would give rise to a current

$$C'=Ne\sqrt{\frac{eXl}{2m}},$$

so that the current arising in this way would be proportional, not to X, but to its square root.

Now, in order that C′ should be of sensible amount in comparison with C, Xel would have to be of the same order of magnitude as $\frac{1}{2}mv^2$, the kinetic energy of an electron, the value of which at 0° C. is about $3{\cdot}6\times10^{-14}$. But the order of magnitude of e is 10^{-20}, so that Xl would have to be comparable with $3{\cdot}6\times10^{6}$. Taking the exceptionally high value of the mean free path found for bismuth by Sir J. J. Thomson, viz. 10^{-4} centimetre, this would require a value of X of the order 10^{11}, or a thousand volts per centimetre, which would be far beyond what could be applied to any known conductor without causing its immediate disintegration.

If the temperature of the conductor is not uniform throughout, the mean kinetic energy of the electrons will be greater in the hotter parts than in the colder ones; and, just as in the case of a hot mass of ordinary gas in contact with a colder one, interchanges of energy will take place between the electrons, tending to equalise the temperature, so that the electrons will carry heat from the hotter to the colder portions of the conductor. We shall presently find that experimental evidence of widely varying character all tends to the conclusion that the conduction of heat in solid electrical conductors is entirely due to this source, and that what we call the temperature of the conductor is, in fact, solely that determined by the mean kinetic energy of the free electrons. In the meantime we shall assume this to be the case, and thereby obtain a relation between the electrical conductivity κ of a metal or other solid, and its thermal conductivity κ'. The latter can be shown by the kinetic theory of gases[1] to be given by the expression

$$\kappa'=\frac{N\beta' lvR}{2}$$

where β' is a *persistence* coefficient. The value of β', calculated on the same assumptions as those made in calculating β, is found to be slightly larger than the latter, but in comparing κ and κ'

[1] Jeans, *Dynamical Theory of Gases*, pp. 259 and 266.

very little error will be made in taking β equal to β', when they will disappear from the result. Writing then β for β', we have

$$\frac{\kappa'}{\kappa}=\frac{N\beta lvR}{2}\cdot\frac{6RT}{N\beta e^2lv}=\frac{3R^2T}{e^2}.$$

The ratio is therefore the same for all metals, and is directly proportional to the absolute temperature. We might calculate this ratio numerically by substituting for R its value as previously given, and taking $e=10^{-20}$, but a more accurate result is obtained by a method employed by Sir J. J. Thomson.[1] The charge e is the same as the charge of an atom of hydrogen, and one electromagnetic unit of electricity liberates 1·2 cubic centimetre of hydrogen at 0° C. (viz. for T = 273°) and at a pressure of one atmosphere (viz. 10^6 dynes). Now, each molecule of hydrogen contains two atoms, so that

$$2{\cdot}4Ne=1.$$

Hence from the equation

$$p=NRT$$

we find that at 0° C.

$$\frac{RT}{e}=\frac{p}{Ne}=10^6\times 2{\cdot}4.$$

Therefore, at 0° C.,

$$\frac{\kappa'}{\kappa}=\frac{3R^2T}{e^2}=\frac{3R^2T^2}{e^2T}=\frac{3(10^6\times 2{\cdot}4)^2}{273}=6{\cdot}3\times 10^{10}.$$

It must be remembered that the equation $p=NRT$ is merely a first approximation,[2] and also that the above value should be multiplied by β'/β, the ratio of the *persistence* coefficients, which is probably somewhat greater than unity. In the present state of knowledge it would, however, I think, be impossible to obtain a closer approximation, and the experimental results of Jaeger and Diesselhorst,[3] obtained at a temperature of 18° C. (or T = 291), suggest that the approximation is a very close one. At 18° C. we should have

$$\frac{\kappa'}{\kappa}=6{\cdot}3\times 10^{10}\times\frac{291}{273}=6{\cdot}70\times 10^{10}.$$

Since this ratio is proportional to the absolute temperature, the temperature coefficient should be ·366 per cent. The results

[1] *Corpuscular Theory of Matter*, p. 56.
[2] See Poynting and Thomson's *Heat*, chap. xi.
[3] *Berlin Academy Sitzungsberichte*, vol. xxxiii., 1899, p. 719.

of their determinations for a number of different metals are given in the accompanying table:—

Metal.	$\frac{\kappa'}{\kappa}$ at 18° C.	Temperature coefficient of $\frac{\kappa'}{\kappa}$. Per cent.
Copper, commerial	$6{\cdot}76 \times 10^{10}$	...
Copper (1), pure	$6{\cdot}75 \times 10^{10}$	0·39
Copper (2), pure	$6{\cdot}71 \times 10^{10}$	0·39
Silver, pure	$6{\cdot}86 \times 10^{10}$	0·37
Gold (1)	$7{\cdot}27 \times 10^{10}$	0·36
Gold (2), pure	$7{\cdot}09 \times 10^{10}$	0·37
Nickel	$6{\cdot}99 \times 10^{10}$	0·39
Zinc (1)	$7{\cdot}05 \times 10^{10}$	0·38
Zinc (2), pure	$6{\cdot}72 \times 10^{10}$	0·38
Cadmium, pure	$7{\cdot}06 \times 10^{10}$	0·37
Lead, pure	$7{\cdot}15 \times 10^{10}$	0·40
Tin, pure	$7{\cdot}35 \times 10^{10}$	0·34
Aluminium	$6{\cdot}36 \times 10^{10}$	0·43
Platinum (1)	$7{\cdot}76 \times 10^{10}$	...
Platinum (2), pure	$7{\cdot}53 \times 10^{10}$	0·46
Palladium	$7{\cdot}54 \times 10^{10}$	0·46
Iron (1)	$8{\cdot}02 \times 10^{10}$	0·43
Iron (2)	$8{\cdot}38 \times 10^{10}$	0·44
Steel	$9{\cdot}03 \times 10^{10}$	0·35
Bismuth	$9{\cdot}64 \times 10^{10}$	0·15
Constantan (60Cu40Ni)	$11{\cdot}06 \times 10^{10}$	0·23
Manganine (84Cu4Ni12Mn)	$9{\cdot}14 \times 10^{10}$	0·27

It will be seen that for many of the metals the observed results are in close accordance with the theoretically calculated ratio and temperature coefficient, but some of them exhibit considerable deviations, which are specially marked in the case of the alloys. With very few exceptions, this decrease in the electrical, relatively to the thermal, conductivity is found to be characteristic of alloys. Investigations made by Dewar and Fleming[1] on the electric resistance of pure metals and alloys show a fundamental distinction between the two. The resistance of pure metals decreases uniformly with decreasing temperature, and would apparently vanish entirely in the neighbourhood of the absolute zero. This is what we should expect if the temperature of a conductor is, as we have assumed, simply that determined by the kinetic energy of the electrons. The atoms must then be regarded as forming a framework of obstacles through which the electrons move. This framework must be continually set into

[1] *Roy. Soc. Proc.*, vol. lx., 1896.

vibration by collisions with the electrons, but its motion must be supposed to be dissipated into radiation so rapidly that the atoms never acquire a motion comparable with temperature motion. When the temperature falls sufficiently, the mean kinetic energy of the electrons will diminish to such an extent that the current C′ will become comparable with the current C, and there will be a sensible deviation from Ohm's law. This deviation will increase until the current C becomes insensible in comparison with C′, and ultimately the mean free path l in the expression

$$C' = Ne\sqrt{\frac{eXl}{2m}}$$

will increase indefinitely, as there will no longer be any collisions between the electrons, but only collisions between the electrons and the comparatively rigid atomic framework; that is to say, C′ will increase indefinitely for a finite value of X, or the resistance will tend to vanish.

In the case of alloys, however, the resistances are found to tend to finite limiting values. If our assumptions are correct, these resistances must be of an entirely different nature from what is defined as electrical resistance, and their nature is indicated by some considerations due to Lord Rayleigh,[1] which he applied to the explanation of the divergence of alloys from the ordinary law of resistance. He observes that in the case of mixtures of metals thermoelectric actions come into play, and give rise to effects which cannot be experimentally distinguished from those due to resistance. This *false resistance*, as it has been called, is simply due to what are known as Peltier E.M.F.'s, arising from the composite character of the metal. The phenomenon known as the *Peltier effect*, from the name of its discoverer, is that when an electric current is made to traverse a composite conductor, such as two or more wires of different metals joined end to end, by the application of an external electric force, an absorption, or a development, of heat is observed at each junction. Whether heat is absorbed or developed at any junction depends on the nature of the metals on opposite sides of it, and on the direction of the current. Since the passage of a current always heats the conductor, the Peltier effect is a differential one; that is to say, what is observed is a decrease or increase, as the case may be, of the normal heating effect. The Peltier E.M.F. is essentially located at the junction of two metals. It is therefore quite distinct from the Volta E.M.F. which was considered at the beginning of Chapter XI., since the latter is mainly located

[1] *Nature*, vol. liv., 1896, p. 154; *Scientific Papers*, vol. iv. p. 232.

at the surface of each conductor, and depends both on the conductor and the surrounding medium. That is to say, the Volta E.M.F., although arising from the charge at the interface, is a function of the surface condition of the conductor, and therefore consists in a superficial distribution of energy. If the temperature of a composite circuit were uniform throughout, the Peltier E.M.F.'s would balance each other when summed up round the circuit, like the Volta E.M.F.'s in a complete metallic circuit. When, however, an externally maintained current is made to traverse such a conductor, this balance is disturbed, as the balance of the Volta E.M.F.'s is disturbed by the introduction of an electrolyte into the circuit, finite differences of temperature being established at the junctions, resulting in the establishment of a resultant thermoelectric E.M.F. round the circuit. This is known as the *Seebeck effect*, and gives rise to a current in the opposite direction to the one by which the differences of temperature are produced, and tending therefore to cool the junctions at which heat is developed by the former, and to heat those at which heat is absorbed. This is the phenomenon on which the construction of the thermoelectric battery, or thermopile, is based, Seebeck's original discovery being that if a circuit were formed of a bar of bismuth and a bar of antimony joined together at the ends, and one junction heated and the other maintained cool, an electric current was produced, flowing from bismuth to antimony across the hot junction, and therefore from antimony to bismuth across the cold one. Since the heat applied at the hot junction forms the sole source of the energy maintaining the current, it necessarily follows, from the principle of the conservation of energy, that, as stated above, the hot junction should be cooled, and the cold junction heated. The application of these principles to an alloy traversed by an electric current is most simply illustrated by supposing the mixed metals to be arranged in layers with the lamination perpendicular to the direction of the current. Differences of temperature proportional to the current will thus be set up, giving rise to a thermoelectric E.M.F. proportional to the current and tending to oppose it, which will produce exactly the same effect as an additional resistance. Since these thermoelectric E.M.F.'s may arise, though in a minor degree, through differences in strain and crystalline arrangement, as well as by differences in molecular constitution, it appears to me that the deviations from the theoretical law, as shown in the table on page 306, may very possibly arise, in part at any rate, from this cause in the case of the so-called pure metals. Impure metals are of course mixtures, and a truly pure metal is practically unattainable, so that this cause

will always operate to a certain extent. The supposition appears to be confirmed by the fact that the metals, other than recognised alloys, which exhibit the greatest increase in resistance, or apparent resistance, are iron and bismuth, in both of which the amount of structure, as shown by microscopic observation, is very considerable.

Since the Peltier effect between two metals is, as stated above, essentially located at the junction, and is independent of the surrounding medium, it must arise from a flow of free electrons from one metal to the other across the junction, provided the movement of electricity in metals is due entirely, as we have supposed, to the motions of these free electrons. Such a flow would arise if there were a difference in the pressure of the free electrons, considered as gases, in the two metals. Since, in order to produce the Peltier effect, this must take place even when they are at the same temperature, the pressure of the electrons in a conductor at a given temperature must depend on the nature of the conductor. Let p_1 and p_2 be the pressures in the two conductors, N_1 and N_2 the numbers of free electrons per unit volume, and v_1 and v_2 the mean velocities of the electrons at the temperature T. Then we have

$$p_1 = N_1RT = \tfrac{1}{3}N_1mv_1^2, \; p_2 = N_2RT = \tfrac{1}{3}N_2mv_2^2.$$

If p_1 is greater than p_2, there will be a flow of electrons from the first metal to the second, charging the latter negatively and the former positively, and this flow will continue until the electric force across the junction, due to these electrifications, balances the effect of the difference in pressure. Let us suppose the transition from the state in one metal to that in the other to take place gradually across a thin layer extending to a short distance on each side of the interface. Let V_1 and V_2 be the potentials on opposite sides of the layer, which we will assume to be perpendicular to the axis of x. Let N be the number of electrons per unit volume, X the electric force, p the pressure of the electrons, and V the potential, at a distance x from one of the boundaries of the layer. Then the electric force parallel to x, acting on the electrons in unit volume at such a point, will be XNe, while the force due to the pressure will be $\dfrac{dp}{dx}$. Then, remembering that $p = \mathrm{NRT}$, and $\mathrm{X} = \dfrac{d\mathrm{V}}{dx}$, and assuming the temperature to be constant within the layer, we have

$$\mathrm{XN}e = \frac{dp}{dx},$$

or

$$\frac{dV}{dx} = \frac{RT}{Ne}\frac{dN}{dx};$$

so that

$$V_1 - V_2 = \frac{RT}{e}\log\frac{N_1}{N_2}.$$

Now, at 0° C. and standard atmosphere pressure (see p. 305),

$$\frac{RT}{e} = \frac{p}{Ne} = 10^6 \times 2{\cdot}4,$$

so that in volts

$$V_1 - V_2 = \frac{10^6 \times 2{\cdot}4}{10^8}\log\frac{N_1}{N_2} = {\cdot}024\log\frac{N_1}{N_2}.$$

In the case of antimony and bismuth, for which the Peltier effect is exceptionally large, $V_1 - V_2$ at 0° C. is about $\frac{1}{30}$th of a volt, which would give $\log(N_1/N_2) = 1{\cdot}33$, or $N_1/N_2 = 3{\cdot}8$. There must, therefore, be about 3·8 times as many electrons in unit volume of antimony as in unit volume of bismuth. As this is an exceptionally large value of the ratio, it will be seen that there is no very large variation in the number of free electrons per unit volume of different metals. This is in accordance with Schuster's results referred to on page 299. We shall see presently that, according to our present working hypotheses, viz., that the temperature of a solid conductor is that due to the electrons, and that the atoms may be considered as forming a stationary framework, two possibly free electrons must be considered as associated with each atom of every solid conductor. The total of two electrons for every atom will only be set free at comparatively high temperatures, which will differ for different metals, so that at a given temperature there will be a difference in the number of free electrons per unit volume, and consequently a Peltier effect, in the case of any possible pair of metals. In the case of bad conductors only a very few of these electrons may be free under ordinary conditions, and a large proportion of the energy required to raise the temperature of such substances may be expended in liberating new electrons. When a metal approaches its fusing point, the atomic energy will largely increase, so that it will no longer be possible to regard its atoms as forming a stationary framework.

In many metals there is a very large increase in the resistance on fusion. In the cases of lead, zinc, and tin, for example, the resistance is practically doubled. This indicates the existence of some other cause, or causes, than a variation in the number of free electrons, for the changes in resistance which accompany

changes in temperature. If this sudden increase in resistance were due to a change in the number of free electrons, it would mean that in the fused metal there were only half as many free electrons per unit volume as in the solid metal at the same temperature. This does not appear probable on the face of it, since there is every reason for supposing, as pointed out above, that the liberation of electrons increases as the temperature rises. Moreover, such an explanation is absolutely negatived by the fact that no sudden change is observed in thermoelectric circuits containing these metals when they pass from the solid to the liquid state, whereas if the number of free electrons were halved, it would give rise to a Peltier effect of about half the magnitude of that observed between bismuth and antimony. We must therefore look elsewhere for the explanation. Now, we found for the specific conductivity of a metal the expression

$$\kappa=\frac{N\beta e^2 lv}{6RT},$$

and in this expression β, e, and R are constants. There is no sudden change in T, for the sudden change in resistance occurs when fusion takes place, and before there is any rise in temperature. The change cannot therefore be due to an increase in T. Neither can it be due to a change in v, for this would involve a change in T, which is proportional to v^2. It can only, therefore, if this expression still applies, be due to a decrease in the mean free path l, so that we have to inquire whether there is any reason to expect such a decrease. Now, observation shows that the assumption that the molecules of the metal may be considered as forming a stationary framework bombarded by the electrons ceases to agree with the experimental facts when the metal is approaching the temperature of fusion. At lower temperatures, as Dulong and Petit have shown, the atomic heats[1] of different solid elements do not vary greatly from the common mean value of about 5·88, except near the fusing point, and, in the case of bad conductors, at temperatures too low for the full number of two electrons per atom to be set free. In the latter case there is evidence of an asymptotic approach to this value as the temperature rises. For example, Weber found[2] the variation of the atomic heat of carbon (diamond) with the temperature to be as given below:—

Temperature	0°	50°	100°	150°	200°	606°	806°	985° C.
Atomic heat	1·12	1·72	2·28	2·81	3·33	5·26	5·36	5·49.

[1] The atomic heat of a body is the product of its specific heat into its atomic mass. See also Poynting and Thomson's *Heat*, p. 86.

[2] *Annalen der Physik*, vol. cliv., 1875, p. 575.

As an example of the former case Pionchon found[1] for the atomic heat of iron at 500° the value 9·84. It follows, therefore, that when a metal is approaching its fusing point the particles which determine the temperature by their mean kinetic energy can no longer be regarded as forming a simple gas with all the particles of the same size, viz. that of the electrons. It must now be regarded as a mixture containing an increasing proportion of the larger metallic atoms. The kinetic theory of gases shows that this will not sensibly affect the temperature, for at temperatures at which any known substances remain in the solid state there is so little interchange of energy with the ether that the system at any moment may without sensible error be treated as a conservative system, in which case the mean kinetic energies of the different types of particles forming the mixture will have equal values, and therefore the temperature will still be that determined by the electrons. The value of l will, however, not remain the same. As we can only determine the general character of this change, which will be sufficient for our purpose, it will not be necessary to consider the complicated formulæ determining the variation of the free path with the changes in composition of the mixture. In the case of a simple gas consisting of uniform hard spherical particles l may be shown to vary inversely as the product na^2, where n is the number of particles in unit volume and a is the radius of a particle. Now we must bear in mind that the addition of metallic atoms, while increasing n, will not increase N in the expression for the conductivity unless the free atoms as well as the electrons are to be considered as carriers of electricity. Therefore the free path will be *decreased*, both by the increase in n, and by the increase in the average value of a, causing a decrease in the conductivity, as observed. If the atoms, as well as the electrons, act as carriers, there will still be a decrease in l, and therefore in the conductivity, though of smaller amount. There are good grounds for believing that the metallic particles which begin to contribute to the kinetic energy as the fusing point is approached, are not single atoms, but comparatively large clusters of atoms, and this, by increasing the value of a, would increase the effect in the direction observed. Moreover, there is reason to believe that these large clusters gradually break up into smaller ones as the temperature increases, and that this breaking up becomes considerable at the moment of fusion. It would appear at first sight that this would tend, by decreasing a, to increase l, instead of decreasing it, and so to diminish the fall in conductivity. This is, however, not the case. The addition of n clusters, each consisting of, say, p atoms, to the mass

[1] *Comptes Rendus*, vol. cvi., 1888, p. 1344.

of moving particles would diminish l more than the addition of n single atoms, but they would diminish it less than the addition of the np separate atoms contained in the clusters. For, consider the two cases of m spheres of radius a per unit of volume, and n spheres of radius b, the amount of matter being the same in each case, so that $ma^3=nb^3$, and let l_a and l_b be the respective free paths. Then we have

$$\frac{l_a}{l_b}=\frac{nb^2}{ma^2}=\frac{a}{b},$$

so that the length of the free path is proportional to the radius of the cluster, and therefore the breaking up of the clusters will further diminish the free path. This formation of clusters in a solid metal and their breaking up on fusion is suggested by Sir J. J. Thomson[1] on the ground of the complex structure of metals revealed by microphotography, which irresistibly suggests the existence in most metals of aggregates of such clusters. This evidence may be considerably reinforced from other directions.

We learn from the kinetic theory of gases[2] that in gases at very high temperatures there are no molecules in combination. As the temperature falls there is an increasing tendency to form into larger and larger clusters. When the clusters become large the analytical method employed ceases to be applicable, but it affords an indication of a very considerable increase in the number and size of the clusters as the point of liquefaction is approached. The analysis shows that this takes place by the molecules remaining for a continually increasing time under the influence of neighbouring molecules, until groups are formed in which the molecules never pass out of each other's sphere of action, but continue to describe orbits about one another. It is evident, from physical considerations, that the continuance of this process beyond the point of liquefaction, until solidification begins, will necessarily result in the groups of molecules most closely united in orbital motion first cohering into solid clusters and aggregates of clusters, and that a large number of these must be rapidly broken up when the process is reversed and the solid is liquefied by fusion. It is, moreover, pointed out by Jeans[3] that it would be difficult to account for the process of conduction of heat between a solid and a gas in contact with it unless the atoms of the former are formed into large clusters which themselves take up the temperature motion from the electrons and convey it to the molecules of the gas.

[1] *Corpuscular Theory of Matter*, p. 71.
[2] See Jean's *Dynamical Theory of Gases*, p. 329.
[3] *Phil. Mag.*, vol. xvii., 1909, p. 794.

If the value of N for a given metal were not a function of the temperature, the thermoelectric E.M.F., E, round a circuit consisting of two metals would be determined by the two equations.

$$V_1 - V_2 = \frac{RT}{e} \log \frac{N_1}{N_2},$$

$$V'_1 - V'_2 = \frac{RT'}{e} \log \frac{N_1}{N_2},$$

where V_1 and V_2 are the potentials on opposite sides of the junction at absolute temperature T, and V'_1 and V'_2 are the potentials on opposite sides of the junction at absolute temperature T'. We should therefore have

$$E = V_1 - V_2 - V'_1 + V'_2 = (T - T')\frac{R}{e} \log \frac{N_1}{N_2},$$

so that the E.M.F. would be directly proportional to the difference between the temperatures of the two junctions. This was at first supposed to be the case, but Cumming observed, in 1823, that on starting at moderately low temperatures, and gradually increasing the mean temperature of the two junctions, while maintaining a constant difference between them, the value of E first decreases to zero, changes sign, that is to say, becomes reversed in direction, and again increases in numerical value. If the temperature of the junctions are equidistant (one above and the other below) from the zero, or *neutral*, point, the value of E is zero. The maximum value of E, for a given difference of temperature, is obtained when one of the junctions is at the neutral point, its sign being determined by whether this is the hotter or the colder junction.

Since the heat supplied to the circuit is the only possible source of the energy maintaining the current, Lord Kelvin pointed out that this reversal showed the presence of other E.M.F.'s than those corresponding to the Peltier effects. He then considered the case of the hot junction being at the neutral point, and therefore the other at a lower temperature. There is then no Peltier E.M.F., and consequently there can be no absorption of heat at the hot junction, and at the cold junction heat is evolved, not absorbed. The principle of the conservation of energy therefore necessitates an absorption of energy in the metals themselves, in consequence of the differences of temperature at their ends. This conclusion he confirmed experimentally, showing that, in an unequally heated conductor, an electric current tends to reduce differences in temperature in some metals, of which copper is an example, while in other metals, of which iron is one, it tends to increase these differences. That is to say, when a current flows from the hot to

the cold end of a copper wire, heat is evolved in the wire, and is absorbed when the current flows in the opposite direction. In an iron wire these phenomena are reversed. Since the electric current in a copper wire tends to decrease the differences in temperature, it behaves like a material fluid having a positive specific heat, while in iron it behaves like a fluid having a negative specific heat. This convection of heat by the current is known as the *Thomson effect*. Lord Kelvin called this apparent specific heat the specific heat of electricity in the corresponding metal.

Consider a straight metallic wire AB with its temperature increasing from A to B. Let p be the pressure of the electrons at a distance x from the end A. Since p increases with the temperature, it will increase with x, and the force per unit area on a plane through x perpendicular to the wire will be $\frac{dp}{dx}$ acting in the direction BA. In order to prevent the electrons from drifting under the action of this force, it must be balanced by an electromotive force X acting in the direction AB, and determined by the equation

$$\mathrm{X}e\mathrm{N} = \frac{dp}{dx},$$

where N is the number of electrons per unit of volume at the point x. Now, $p = \mathrm{NRT}$, so that

$$\mathrm{X}e = \frac{\mathrm{R}}{\mathrm{N}}\,\frac{d(\mathrm{NT})}{dx}.$$

An electron travelling from the point $x + dx$ to x will therefore absorb from the metal an amount of heat

$$\mathrm{X}e\,dx = \frac{\mathrm{R}}{\mathrm{N}}\,\frac{d(\mathrm{NT})}{dx}dx,$$

the heat being expressed in mechanical measure, that is to say, as energy.

Now, an electron having the kinetic energy $\frac{3}{2}\mathrm{RT}$ at x, would have had the kinetic energy $\frac{3}{2}\mathrm{R}\left(\mathrm{T} + \frac{d\mathrm{T}}{dx}dx\right)$ at the point $x + dx$. That is to say, in travelling from $x + dx$ to x, it will communicate to the wire between these points an amount of heat energy equal to $\frac{3}{2}\mathrm{R}\frac{d\mathrm{T}}{dx}dx$. The total amount of heat absorbed by the wire from an electron travelling from $x + dx$ to x will therefore be

$$\left\{\frac{3}{2}\mathrm{R}\frac{d\mathrm{T}}{dx} - \frac{\mathrm{R}}{\mathrm{N}}\,\frac{d(\mathrm{NT})}{dx}\right\}dx = \mathrm{R}\left\{\frac{3}{2} - \frac{1}{\mathrm{N}}\,\frac{d(\mathrm{NT})}{d\mathrm{T}}\right\}d\mathrm{T}.$$

Now, if the direction of the current, C, is from A to B, the number of electrons crossing a section of unit area in a second in the direction BA will be C/e, and the heat energy absorbed from them by the wire between points where the temperatures are T and $T+dT$ respectively will be

$$\frac{CR}{e}\left\{\frac{3}{2}-\frac{1}{N}\frac{d(NT)}{dT}\right\}dT=-CsdT,$$

where s is the specific heat of electricity in the metal, the negative sign being taken, since the direction of the current is from the colder to the hotter end of the wire, that is to say, the direction in which T is decreasing. We have therefore

$$s=-\frac{R}{e}\left\{\frac{3}{2}-\frac{1}{N}\frac{d(NT)}{dT}\right\}=-\frac{R}{e}\left\{\frac{3}{2}-\frac{T}{N}\frac{dN}{dT}-1\right\}$$

$$=-\frac{R}{2e}\left\{1-2T\frac{d\log N}{dT}\right\}=\frac{RT}{e}\frac{d}{dT}\log NT^{-\frac{1}{2}}.$$

The term $R/2e$ is a constant for all conductors, and therefore will not affect the thermoelectric E.M.F. round a circuit of two metals, as this involves only the difference in the values of s for the two metals; but Sir J. J. Thomson points out [1] that it will affect the amount of heat developed in the conductor, and that unless it very nearly balanced the term $2Td\log N/dT$, the development, or absorption, of heat by a current flowing through an unequally heated conductor, would be far greater than the amount actually observed. He also points out that in bismuth, for which the heat effect is largest, as far as our present knowledge extends, it is only about a third of the value of $R/2e$, and as this is much greater than is observed in the case of any other metal, s may be taken as being always small compared with $R/2e$. It follows therefore, that the equation

$$s=-\frac{R}{2e}\left\{1-2T\frac{d\log N}{dT}\right\}$$

is approximately equivalent to

$$\log N=\tfrac{1}{2}\log T+\text{a constant},$$

so that N will be approximately proportional to the square root of the absolute temperature, the variation being a little more rapid than this when s is positive, and a little less rapid when s is negative. Since the electrical conductivity has been shown to be

[1] *Corpuscular Theory of Matter*, p. 79.

proportional to Nbv/T, and v is proportional to $\sqrt{T}$, while N varies approximately according to the same law, it follows that the conductivity must be approximately proportional to the length of the mean free path of the electrons. In the case of many pure metals the conductivity is found to vary approximately as the reciprocal of the absolute temperature, and it follows, therefore, that the length of the mean free path in these metals must vary according to the same law; and, as Thomson points out, such a rapid variation of the mean free path with the temperature would be impossible if the structure of a metal were analogous to that of a gas compressed so that the distances between the molecules were all diminished in the same proportion, that is to say, if metals did not consist largely of aggregates of clusters.

It was mentioned earlier in this chapter that thermoelectric E.M.F.'s may arise through variations in mechanical strain at different portions of a metallic circuit. This was discovered experimentally by Lord Kelvin,[1] who found that when a weight was applied so as to produce a state of strain in a portion of a wire forming a complete circuit, and the two junctions of strained and unstrained portions were maintained at different temperatures, a current was produced the direction of which was from the unstrained to the strained portion across the hot junction in iron and steel wires, while in copper wires the current was in the opposite direction.

The experiments were repeated by Le Roux,[2] who in each case obtained currents in the opposite direction to those observed by Lord Kelvin. This discrepancy was accounted for by the present writer,[3] who found that when the strain was gradually increased from a small value up to the breaking point of the wire, the E.M.F. first increased to a maximum, then began to diminish, and finally, just before the breaking strain was reached, reversed its direction. When the strain was put on suddenly it was found, in iron and steel wires, to give rise to temporary oscillations in the E.M.F. above and below the value corresponding to the strain, indicating a tendency for the structure of the wire to return, after overstrain, towards its initial condition, just as a stretched spring oscillates when set free.

These results obviously lead to the conclusion that the changes in the structure of the metal, due to variations in the strain, give rise to variations in the value of N, and the complexity of the phenomena suggests changes also in the value of l, arising, in

[1] "Electrodynamic Qualities of Metals," *Phil. Trans.*, 1856.
[2] *Annales de Chimie et Physique*, vol. x., 1867, p. 201.
[3] *Phil. Mag.*, May 1878.

all probability, from the disintegration and re-formation of clusters of molecules.

The fact that the free electrons move rather more rapidly down a temperature gradient than up it, which is the basis of the Thomson effect, leads to the conclusion that an electromotive force must arise from the interaction of a magnetic field and a temperature gradient, its direction being at right angles to both these vector quantities.[1] The existence of such an electromotive force has been experimentally detected, as also of the resulting converse phenomenon, by von Ettingshausen and Nernst.[2]

Our theory also affords a simple explanation of very interesting photoelectric phenomena, the most important of which is the result of exposing selenium to light. This was first observed by May, an assistant to Willoughby Smith,[3] who noticed that the electric resistance of a piece of selenium carrying an electric current was diminished by exposure to light. The phenomenon was then very fully investigated by Professor W. G. Adams and R. E. Day,[4] after the former had shown that an electromotive force, as well as a diminution of resistance, resulted from the exposure of selenium to light. The results were such as could only be due, from our present point of view, to the liberation of free electrons by the action of light. The results of Professor G. M. Minchin's very interesting researches on photoelectric cells[5] can be similarly accounted for. We shall see in Chapter XXI. that exposure of a conductor to any form of radiation of comparatively short wave-length may be expected to give rise to the liberation of additional free electrons, and that a greater effect is to be anticipated from a given amount of radiant energy in the form of light-waves than when in the form of heat-waves, owing to the smaller wave-lengths in the former case.[6]

It was found by Hall[7] that when a metallic conductor is placed in a magnetic field the lines of flow of an electric current are

[1] The magnitude of this electromotive force must depend on the vector product of the two quantities by which it is determined, and this will also apply to the Hall effect.

[2] See Riecke's *Experimental Physik*, vol. ii. p. 327.

[3] *Nature*, vol. vii., 1873, p. 303.

[4] See *Roy. Soc. Proc.* for 1876 and 1877, and *Phil. Trans.* for 1877.

[5] See *British Association Report*, 1880, p. 468.

[6] Since the above was written I find that the same explanation of the action of light on selenium has been suggested by Professor Nagaoka (*Phil. Mag.*, vol. vii., 1904, p. 455), who further suggests that the effect of electric waves incident on a coherer may be accounted for in the same manner, bearing in mind that the resonance effect will be much greater in the case of these long waves than in the case of the shorter waves corresponding to visible light.

[7] *American Journal of Mathematics*, vol. ii., 1879, p. 287; or *Phil. Mag.*, vol. ix., 1880, p. 225.

distorted in such a manner as would arise from the presence of an electromotive force acting at right angles to the electromotive force producing the current and to the direction of the magnetic force. Suppose, for example, an electromotive force to produce a current from east to west in a thin plane sheet of metal lying in a horizontal plane. The effect of a uniform magnetic field with the magnetic force acting from north to south will be to cause a distortion of the flow, such as would arise from a small electromotive force acting vertically. In the case of some metals, for example, bismuth and silver, this electromotive force would be directed downwards; while in the case of others, as, for example, iron, cobalt, and tellurium, it would be directed upwards. This electromotive force is not always proportional to the magnetic force, and in the case of some alloys it is found that its direction can be reversed by varying the amount, without changing the direction, of the magnetic force.

If H is the downward acting magnetic force, u the average drift of the negative electrons from west to east, corresponding to the current flowing from east to west, then each electron will be acted on by a vertical upward force $\mathsf{H}eu$ corresponding to a vertical downward electromotive force, as observed in the case of bismuth and silver. The observed complexity of the Hall effect shows that this cannot be a complete explanation. Now, this explanation only takes account of the effect of the magnetic force on an electron during the time of its free path, and leaves out of consideration its effect on the collisions between the free electrons and the molecules of the body. It is easily seen that this is illegitimate if the molecules of the conductor have a sensible magnetic moment. Moreover, the Hall effect is an extremely minute one, and we have seen that all bodies are more or less magnetic. Suppose, first, that the conductor is paramagnetic. Its molecules will then arrange themselves under the action of the magnetic field with their south poles directed more or less upwards and their north poles directed more or less downwards. Close to one of these molecules, and in the region between its poles, the magnetic force due to the molecule will tend to oppose that of the external magnetic field, and very close to the molecule it may greatly exceed that of the external field. An electron, when in collision with a molecule, would then be deflected in a direction opposite to the deflection produced by the external field during its free path. The expression for the Hall effect would therefore consist of two terms of opposite sign, one arising from the deflections of the electrons by the external field during their free paths, and the other from their deflections by the molecules during their collisions. If the substance of the conductor were

diamagnetic, the two terms would, on the other hand, be of the same sign, the magnetic action of the molecules during the collisions being in the same direction as that of the external field. Further complexities would evidently result from the varying aggregations of the molecules into clusters of which we have found evidence.

All recent investigation of the radiation of solids may be considered as founded on the basis of the two experimental laws known as the laws of Stefan[1] and of Wien[2] respectively, both of which were discovered in the year 1879. Stefan's law states that the rate of radiation is proportional to the fourth power of the absolute temperature. Wien's law is that if a body at the absolute surface temperature T is emitting its full radiation, then the maximum wave-length λ_m of the radiation is determined by the equation

$$\lambda_m T = \text{constant}.$$

Wien also showed that the energy E_m, which is radiated through a small range of wave-length $d\lambda$ at this maximum, is determined by the equation

$$E_m d\lambda = \text{constant} \times T^5.$$

Attempts have been made to establish both these laws on a theoretical thermodynamic basis, but the arguments involve assumptions which cannot be considered justifiable. Jeans has shown, however, that their approximate truth may be demonstrated on the basis of the electron theory developed in this chapter, which proves them, moreover, to be only approximations, and not exact laws (see Appendix M). Another important law of radiation, known as Kirchhoff's law, is there shown to follow from the electron theory. The law states that if A is the coefficient of absorption of radiation incident on any body, the intensity of the stream of issuing radiation will be A times the intensity of the stream issuing from a black body at the same temperature. The law holds for the radiation of any given wave-length, and therefore also for the total radiation. Kirchhoff's proof was based on Carnot's principle, a procedure which Jeans has shown to be illegitimate in the case of radiation problems.[3] The demonstration of Carnot's principle is usually based on an appeal to experience in assuming the tendency to equalisation of temperature, and in the course of the argument it has to be assumed that the energy of the working substance is a function of only two indepen-

[1] *Wien. Akad. Ber.*, vol. lxxix., 1879, p. 391.

[2] *Congrès International de Physique*, vol. ii. p. 23. See also Poynting and Thomson's *Heat*, p. 248.

[3] *Roy. Soc. Proc.*, vol. lxxvi., A, 1905, p. 296.

dent variables, as, for example, the temperature and the density. This is not true where ether is the working substance, for the ether energy consists of the sum of a number of vibrations of different wave-lengths, and the number of vibrations to be included in this sum will vary with the nature as well as the temperature of the matter with which the ether is in communication. If we attempt to avoid this difficulty by considering only the cases in which all the matter is the same, we are confronted with a still more fundamental one. The piston of the engine would form a moving boundary for the ether, and it can be shown that the result of this would be to change the frequency of its vibrations, and to rapidly transfer energy, at every stroke of the pump, to the highest frequency vibrations which, as a matter of experience, do not come into play at all. Instead of basing Carnot's principle on the usual appeal to experience, it might be derived mathematically from statistical mechanics. It would then, however, only be applicable to conservative systems, that is to say, to systems in which the final state has been attained. A system of matter and ether does not fulfil these conditions, and would only do so if the transfer of energy between matter and ether took place with extreme rapidity. This is not, however, the case, except for the vibrations of low frequency. For those of high frequency the transfer is extremely slow.[1]

An important consequence of the establishment of Kirchhoff's law on the basis of purely electron theory is pointed out by Jeans.[2] The law is thus shown to be true quite independently of whether the partition of energy between ether and matter has or has not attained the final condition of equilibrium, and is, in fact, entirely independent of thermodynamic considerations of all kinds. It is, therefore, illegitimate to draw any thermodynamical inferences from the fact that Kirchhoff's law is observed to be true in nature.

We have now to consider the researches of Hagen and Rubens,[3] previously referred to, and how their results, in connection with the formula expressing the variation of electric conductivity with frequency, give us information as to the value of N. Jeans points out that, since the velocity of an electron is about 10^7, and the radius of an atom about 10^{-8} cm., the time of collision is probably about 10^{-15} sec. For light of wave-length 4 μ, where μ represents a *micromillimetre*, that is to say, a thousandth of a millimetre, the time of vibration would therefore be about thirteen times the time of collision; and he considers, therefore, that it should be legitimate to use the formula for light wave-lengths 4 μ, or greater, but probably not for wave-lengths much less than this.

[1] See Appendix M.

[2] *Phil. Mag.*, vol. xvii., 1909, p. 785.

[3] *Ibid.*, vol. vii., 1904, p. 157.

Earlier investigations made by Hagen and Rubens had shown that the relation between the transparencies of metals and their electric conductivities approached more nearly, as the wave-length increased, to those required by Maxwell's electromagnetic theory of light, in its original form in which no account was taken of the molecules and their period of vibration. In the paper referred to above these investigations were extended to wave-lengths of periods varying from 4 μ to 25·5 μ, with the anticipated result of demonstrating a much closer agreement between the observed and the theoretical values of the transparencies, the latter being calculated from the observed conductivities. The direct measurement of transparency for these longer waves would be extremely difficult, if not impossible, to carry out with accuracy, owing to the intensity of radiation necessary, and to the fact that there are no known substances which are at the same time sufficiently transparent to heat rays and suitable to act as mechanical supports for the extremely thin layers of metal which must be employed. It is, therefore, much easier to determine the intensity of the radiation entering a metal by measuring either the reflecting power or the emissive power.

If R represents the reflecting power of a metal for normal incidence, expressed as a percentage of the incident radiation, the intensity I of the radiation entering the metal will be

$$I = (100 - R).$$

In those parts of the spectrum in which the energy is sufficient for exact measurements, and R differs sufficiently from 100, the measurement of R affords the best means of determining I, and this is the method employed by the authors for wave-lengths up to 12 μ. For greater wave-lengths these conditions are not adequately fulfilled, and I is then determined by measuring the power of emission. If E_m be the emission of an opaque polished metal surface, and E_0 that of a black body of the same temperature, both for the same wave-length, then $I = E_m/E_0$, by Kirchhoff's law.

We shall only need to consider, for our present purpose, the results obtained by measurements of R. Now, it is shown in a note to the paper that if (κ) is the value in absolute electrostatic units of the conductivity for steady currents, which we have taken as κ, when expressed in electromagnetic units, and τ is the time, in seconds, of a complete vibration, then, on a theory which neglects the variation of conductivity with frequency,

$$R = 100 - \frac{200}{\sqrt{\tau(\kappa)}}\cdot$$

The measurements in the paper are expressed in terms of a conductivity κ' defined as the resistance in ohms of a conductor 1 metre long and 1 square millimetre in sectional area. Expressed in terms of this unit, and introducing the wave-length λ in place of the period τ, the above equation becomes

$$R = 100 - \frac{36 \cdot 5}{\sqrt{\lambda \kappa'}},$$

or

$$(100 - R)\sqrt{\kappa'} = \frac{36 \cdot 5}{\sqrt{\lambda}}.$$

Let k' be the value of the conductivity for currents of frequency p, expressed in terms of the same unit as κ', and let

$$(100 - R)\sqrt{\kappa} = C_\lambda \text{ and } \frac{36 \cdot 5}{\sqrt{\lambda}} = C'_\lambda.$$

Then C_λ would be equal to C'_λ, on the theory which neglects the variation of the conductivity with the frequency; but if this be taken into account we shall have

$$(100 - R)\sqrt{k'} = C'_\lambda,$$

and therefore

$$\left(\frac{C_\lambda}{C_\lambda}\right)^2 = \frac{\kappa'}{k'} = \frac{\kappa}{k} = 1 + \frac{\kappa^2 p^2 m^2}{N^2 e^4},$$

provided λ is not less than four micromillimetres.

This equation, however, neglects resonance effects, and Jeans points out that this is shown to be unjustifiable by the fact that in many of Hagen and Rubens' experiments the value of C_λ/C'_λ was found to be less than unity. To avoid this difficulty, in so far as it can be avoided, he follows a method suggested by Schuster,[1] and replaces this equation by the modified one

$$\left(\frac{C_\lambda}{C'_\lambda}\right)^2 = 1 + \frac{\kappa^2 p^2 m^2}{N^2 e^4} - S,$$

where S is a term required to allow for the effects of resonance, and which can be shown, from the principle of the conservation of energy, to be necessarily positive. It follows, therefore, that the true value of N is necessarily less than the value calculated from the unmodified equation. None of the metals appear to be sufficiently regular in their optical behaviour to justify the entire

[1] *Phil. Mag.*, vol. vii., 1904, p. 154.

omission of the correcting term S. One of the most regular is platinum, for which Hagen and Rubens found

$$\left(\frac{C_\lambda}{C'_\lambda}\right) = 2\cdot 0, \text{ and } 1\cdot 17, \text{ for } \lambda = 4\,\mu \text{ and } 8\,\mu, \text{ respectively.}$$

The corresponding upper limits for N are $2\cdot 3 \times 10^{23}$ and $2\cdot 7 \times 10^{23}$, respectively. Now, ·75 calorie[1] is required to raise the temperature of $2\cdot 3 \times 10^{23}$ electrons through 1° C., while ·69 calorie only is required to raise a cubic centimetre of platinum through the same amount. The greatest value of N compatible with the known specific heat of platinum is $2\cdot 1 \times 10^{23}$.

In the case of gold, which is less regular in its optical behaviour, the upper limit of N, determined by our equation, is 6×10^{23}, while the greatest value compatible with the known specific heat of gold is $1\cdot 9 \times 10^{23}$.

In both these cases the upper limit determined by the equation is greater than that already fixed by the specific heat of the metal, and this holds good for all the metals examined, except steel.

In the case of steel, the upper limit for N, determined by the equation for $\lambda = 8\,\mu$, is $\cdot 96 \times 10^{23}$, while the specific heat-limit is as great as 3×10^{23}. Jeans suggests that the former limit is probably untrustworthy in this case, since the theory takes no account of the magnetic properties of the substances considered, the magnetic permeability being assumed to be equal to unity in every case. It appears quite probable that this may be illegitimate for waves of such long period in the case of steel.

The foregoing results are in general agreement with Schuster's conclusions, which were based on an application of Drude's equations to visible light. If n is the number of atoms in a cubic centimetre of a metal, Schuster's law is expressed by the relation

$$N = qn,$$

where q lies between 1 and 3.

Assuming, as we have done, that the energy of the electrons is that which corresponds to the temperature of the body, Dulong and Petit's law, that the *atomic heat* is nearly the same for most solid elements, sets an upper limit of 2·2 to the value of q for all metals, and a still lower one in the case of metals whose atomic

[1] The calorie may be defined with sufficient exactness for our present purpose as the amount of heat required to raise a gramme of water through a degree Centigrade. Since the specific heat of water varies with the temperature, it is necessary for the maximum of exactitude to state the temperature limits employed in determining the unit. See Poynting and Thomson's *Heat*, p. 65.

heat is less than 6·5. In the case of platinum, for example, the atomic heat is 6·29, which would give an upper limit for q of 2·1, as compared with 2·3 derived from the equation connecting k and p, and 1·91 derived by Schuster's method.

Jeans points out that when the limiting value for q is found to be close to the limit allowed in this way by the specific heat, the inference is that almost all the heat energy of the substance may reside in the free electrons, as was assumed on page 304 to be the case, and that the atoms must then form an almost stationary network of obstacles through which the electrons move. We must now consider further evidence of the validity of this assumption.

The conclusion is arrived at in Appendix M that the energy in a solid conductor is equally divided amongst the vibrations of great wave-length, both in a cavity within it and throughout its substance, so that each vibration has the energy appropriate to an absolute temperature T determined by the kinetic energy of the free electrons. There are, as I shall now proceed to show, two pieces of evidence which justify us in identifying this temperature T with what we call the temperature of the matter.

The first of these is derived from Lummer and Pringsheim's observations of the energy of radiation of great wave-length;[1] the second from Richardson and Brown's observations of the energy of the electrons which escape from hot metals.[2] Lummer and Pringsheim's paper, referred to below, forms a continuation of their earlier researches on the accuracy of the law of Wien, and Planck's modification of it,[3] and contains a very complete list of references to previous work on the subject.

Wien's formula is found to agree well with observation for very short wave-lengths, for which it does not differ sensibly from Planck's modification. As the wave-length is gradually increased, the divergence from Wien's law increases, and for a considerable range the results accord fairly closely with Planck's more complex formula. When the wave-length is still further increased, Planck's formula ceases to be applicable, and Lummer and Pringsheim replace it by one still more complex. For the comparatively great wave-lengths in the ultra-red portion of the spectrum, Planck's formula is easily seen to lead to the conclusion that the radiation of a given wave-length is proportional to the absolute temperature of the substance,[4] and the same holds good with respect to Lummer and Pringsheim's

[1] *Verhandl. d. deutscher phys. Gesellschaft*, vol. ii., 1900, p. 163.
[2] *Phil. Mag.*, vol. xvi., 1908, p. 353.
[3] See Poynting and Thomson's *Heat*, p. 249.
[4] *Ibid.*, p. 342.

formula for greater wave-lengths than those for which Planck's expression can be employed. The observations, therefore, lead to the result theoretically obtained in Appendix M, that the energy of a substance is equally divided amongst the vibrations of great wave-length, but that it is now the energy appropriate to the absolute temperature of the substance. This temperature must therefore be identical with that determined by the kinetic energy of the electrons. The ultra-red spectra were obtained by means of prisms of fluorspar and sylvine, the latter being employed for the longest waves, for which its absorption is relatively small. The energy of the radiation was measured by means of a bolometer.[1] The radiating source consisted of a small hole in the side of an enclosure maintained at a constant temperature, the enclosure being formed of different materials according to the temperature.

The method of investigation adopted by Richardson and Brown consisted in measuring the rate at which a metal plate became negatively charged when a parallel sheet of platinum was maintained at a sufficiently high temperature to give rise to the emission of electrons. The results of the experiments led to the conclusion that the velocities of these emitted electrons were distributed in accordance with Maxwell's law. Those electrons within the metal which are free, in the sense of the kinetic theory of gases, must therefore be distributed according to the same law, since the only difference between those emitted and those remaining within the metal must be due to the difference in their potential energies. Now we know from the kinetic theory of gases that when two regions of the same gas at the same temperature are compared, the regions being such that the mean potential energy of the molecules is different in the two regions, the mean translational energy is the same in both, and is distributed according to the same law. The only effect of the difference in the potential energies is therefore a difference in the concentration of the molecules in the two regions. The mean translational energy of the internal and external electrons must therefore be identical in amount and distribution. The average translational energy, and the law of distribution of energy, of the free electrons within a metal are therefore the same as for the molecules of a gas at the same temperature as that of the metal. The tempera-

[1] The bolometer, invented by Professor Langley, consists of a Wheatstone bridge, one arm of which is formed of a thin sheet of metal, which can be exposed to radiation by sliding back a protecting shutter. The resulting rise in temperature, and consequently the heat energy received by the strip, can then be determined with great delicacy from the observed change in its electrical resistance.

ture determined by the electrons is therefore the same as that which we call the temperature of the metal.

There are, then, two ways in which the temperature of a metal or other solid conductor may be defined, viz. by (1) the mean kinetic energy of the free electrons in its interior; or (2) the mean energy of the vibrations of great wave-length in its interior. The first of these explains the tendency of the temperatures of two bodies to become equalised by conduction; the second explains the tendency of their temperatures to become equalised by radiation.

If, as we have seen reason to believe, the molecules of solid bodies are aggregated into clusters, each containing a considerable number of molecules, we shall now proceed to show that a third definition of the temperature of a solid may be added to the foregoing, viz. it may be defined by the mean kinetic energy of its molecular clusters. In a gas the temperature is defined in only one way, viz. by the mean kinetic energy of translation of its molecules; but for reasons which will immediately appear, the temperature of a solid cannot be similarly defined.

In nearly all metals and other solid elements the atomic heat has, approximately, the value 5·88, which is the value required if each atom had associated with it an amount of energy equal to 3RT, that is to say, twice the translational energy of a free electron, or molecule of gas, at the temperature T. If the energy of motion of the atom is determined by this temperature, each atom must, in virtue of this motion, have associated with it an amount of energy exactly equal to 3RT, half of this being contributed by its average kinetic energy, $\frac{3}{2}$RT, and half by its average potential energy, which is of the same value. It is, however, hardly possible to imagine the atoms moving freely as regards translational motion without simultaneous motions of rotation, and the energy of this rotation, by the law of equipartition of energy, would also amount to 3RT for each atom, if determined by the temperature T. Further, every atom has associated with it a number of free electrons averaging, as we have seen, about two to each atom. This will make a further contribution of 3RT to the energy of each atom, making 9RT in all. Now, the value required by Dulong and Petit's law of specific heat is 3RT in all cases, and this constancy indicates that it is to be attributed to a similar origin in all substances. The only possible conclusion appears to be that it is due entirely to the motion of electrons, and that two free electrons, or rather two electrons which may possibly become free at sufficiently high temperatures, are associated with each atom.

If this view is accepted, the rejection of the contribution 6RT from the atoms means that their energy of motion is negligibly

small in comparison with what it would be if determined by the temperature; that is to say, the atoms must, as previously suggested, form a practically stationary framework of obstacles through which the electrons move, and in order that this may be the case, the vibrations continually set up in the atomic framework by collisions with the electrons, must be dissipated into radiation so rapidly that the atoms never acquire an energy of motion comparable with that corresponding to the temperature.

Although these considerations preclude the possibility of the temperature of a solid being determined, like that of a gas, by the translational energy of the atoms, they would not prevent its being determined by the mean kinetic energy of its molecular clusters, since the energy contribution of 3RT or 6RT from each cluster would be insufficient to add perceptibly to the specific heat.

The conclusion that the temperature of a solid can not be determined by the translational energy of its atoms is confirmed by an investigation by M. Brillouin[1] in which the same result is arrived at by means of a purely thermodynamic argument, without reference to the electron theory.

[1] *Ann. Chim. Phys.*, vol. xviii., 1909, p. 387.

CHAPTER XV.

GENERAL PHENOMENA OF RADIOACTIVITY.

THE fact that Röntgen rays were found to be capable of exciting phosphorescence to a very marked degree in a large number of substances naturally led to investigations to determine whether this phenomenon was generally accompanied by Röntgen radiation. Early in 1896 Henry observed that sulphide of zinc, which had been rendered phosphorescent by exposure to the light of the sun or of burning magnesium, was capable of acting upon a photographic plate through a screen of black paper or thin aluminium foil. A little later in the same year Becquerel obtained a similar action with the double sulphate of uranium and potassium. In order to determine whether this effect were due to phosphorescence persisting from a former exposure to light, Becquerel first dissolved some crystallised uranium nitrate in water in a dark room, recrystallised it, and then found that it acted strongly on a photographic plate. He also obtained a similar effect from the solution of uranium nitrate, which is not phosphorescent. Becquerel therefore concluded that the photographic effect was not due to phosphorescence, but was due to properties belonging to the metal uranium. This conclusion was confirmed by further experiments, in which both metallic uranium and salts of uranium were found to retain these properties after being kept in the dark, some of them in lead boxes, for more than a year. He also found that this radiation, sometimes referred to as *Becquerel rays*, caused ionisation in gases exposed to it.

We shall see presently that other substances, in addition to uranium and its compounds, have been found to possess this same property of spontaneously emitting radiation capable of producing definite photographic and electrical effects, together with others which will be referred to later. Such substances are said to be radioactive, and the phenomenon itself is known as radioactivity.

Rutherford, in 1899, published the results of an extensive investigation of the radiation from uranium and its compounds, in the course of which he discovered that it was composed of two

distinct portions, which he called alpha-rays and beta-rays, respectively. The latter were found to have a high penetrative power, and were soon afterwards shown to be capable of deflection by magnetic and electric fields, and further investigation demonstrated conclusively that they consisted, like the cathode rays, of streams of negative electrons moving at speeds comparable with, though less than, the velocity of light. The alpha-rays, on the other hand, were found to be absorbed by comparatively thin layers of matter, and did not at first appear to be deviable by either magnetic or electric fields. Their ionising power was much greater than that of the beta-rays. Later on Rutherford succeeded, by the adoption of extremely delicate methods of observation, in showing that the alpha-rays were capable of deviation by electric and magnetic fields, but to a much smaller extent than the beta-rays. He also obtained further results leading to the conclusion that they consisted of comparatively heavy particles moving at comparatively low speeds, carrying positive charges, and consisting, in all probability, of molecules either of hydrogen or helium. In the course of these later investigations, a third class of rays, to which the name gamma-rays was given, was discovered. These rays are extremely penetrating, and have not been found capable of any deviation by either magnetic or electric fields.

In 1898 it was discovered independently by Schmidt and by Madame Curie that thorium and its salts possessed radioactive properties similar in their general character to those exhibited by uranium. Detailed investigations by Owens, Rutherford, Soddy and others showed that the radioactive phenomena exhibited by thorium compounds were of a much more complex character than those observed in the case of uranium.

It has been shown that the alpha-rays and the beta-rays emanating from uranium are due to two distinct constituents of this metal, and the phenomena observed in the case of the thorium emanations suggest a much greater complexity in the composition of the metal thorium. Becquerel's discoveries led Monsieur and Madame Curie to make a systematic and extensive examination of chemical elements and compounds, and of minerals. It was only in the case of the minerals that successful results were obtained. All the minerals found to possess radioactive properties contained either uranium or thorium, but some of them proved to be even more radioactive than the pure metals. This led to a systematic examination of the constituents of pitchblende, the matrix in which these minerals occurred, and resulted in the discovery of three distinct and strongly radioactive metals: radium, discovered by Monsieur and Madame Curie and Monsieur Bemont;

polonium, discovered by Monsieur and Madame Curie; and actinium, discovered by Monsieur Debierne. Radium is chemically very closely allied to the metal barium, and when barium is extracted from pitchblende, the radium remains associated with it, but the chloride of radium being more soluble in water, alcohol, or hydrochloric acid than that of barium, it can be separated from the latter by fractional crystallisation. Several tons of pitchblende, or of the residues remaining after the extraction of uranium, have to be treated, in order to obtain a few decigrammes of radium. Madame Curie estimates that the radioactivity of radium is about a million times as great as that of uranium, but owing to the ratio being so high, it is extremely difficult to obtain anything approaching an accurate determination, and, moreover, the ratio will vary to a considerable extent according to which of the three types of radiation is selected for the comparison. As exemplifying the intense activity of radium, Monsieur and Madame Curie have prepared specimens which, when enclosed in a leaden tube half a centimetre in thickness, will discharge an electroscope more readily than can be effected by bringing uranium without any covering close up to the electroscope. It has also been found that the velocity of some of the particles projected from radium is greater than it has yet been possible to produce by electrical means, and, owing to their greater velocities, their penetrating power is much greater than that of the cathode rays.

Polonium is found in pitchblende in association with bismuth, and differs from all other known radioactive substances in emitting only alpha-rays. Actinium occurs in association with thorium, to which it appears to be closely allied in chemical properties, though many thousand times more active. Radium requires very careful handling, as, if the skin is exposed to its radiation for any length of time, even through fairly thick metal coverings, painful and long-continuing sores are produced.

Secondary rays are produced by the impact of radioactive rays on solid obstacles. Those obtained from the alpha-rays are very feeble, but strong secondary rays are obtained from the beta-rays; and so strong are those arising from the gamma-rays emitted by radium, that their effect on a photographic plate is in many cases found to be increased by the interposition of a metallic screen.

In 1893 P. Curie and A. Laborde made the remarkable discovery that a radium compound maintains itself continuously at a temperature several degrees above that of the surrounding atmosphere, and after making numerous measurements, they arrived at the conclusion that the heat emitted by a mass of radium in an hour would be sufficient to raise the temperature of

its own mass of water from freezing point to boiling point, a result which was afterwards confirmed by other observers.

There is hardly room for doubt, in view of the experimental facts, that this heat is derived, not from external sources, but from an actual disintegration of the atoms of radium. The old theory of the stability of the atom could not well be retained, in view of phenomena which have already been considered, and some of the more recent investigations into the phenomena of radioactivity afford still more conclusive evidence of the instability of the atom. Some of the results obtained, indeed, appear practically to amount to a demonstration of the possibility of the transmutation of the so-called chemical elements, so firmly believed in by the early alchemists who, in spite of their many vagaries, were the pioneers of modern chemistry.

Radioactivity is certainly a property, not of the molecules of the radioactive substance, but of its atoms. The rate of emission of the rays is determined by the amount of the radioactive element only, quite independently of what may be its combination with inert substances. Moreover, it is found that wide variations in temperature have no effect on the rate of emission, nor can it be altered by any known chemical or physical means. It does not appear to be at all probable that the projected particles can acquire the enormous velocities which they have been shown to possess, by means of forces called into play at the moment of disintegration. If, however, we adopt the theory which has been developed by Sir J. J. Thomson, Sir Joseph Larmor, and many other physicists, that the chemical atom is a complicated structure built up of large numbers of charged particles in rapid orbital or oscillatory motion, we have only to assume that the initial velocities of the positively and negatively charged escaping particles are approximately those which they possessed while still forming part of an atom. Experimental evidence shows that the alpha-particles resulting from the disintegration of radium would be practically all absorbed after traversing a layer of radium not more than a hundredth of a millimetre in thickness, so that the greater proportion of the alpha-particles, and some at least of the others, resulting from the disintegration of the internal atoms of a mass of radium, would be stopped within the mass itself, and their energy would therefore appear in the form of heat, as was pointed out by Sir Oliver Lodge in 1903. This violent expulsion of portions of an atom must necessarily give rise to very intense electrical disturbances within the atom, and the residual portions will rearrange themselves under the new conditions in such a manner as to form a system which is at least temporarily stable. This process will be accompanied by a further decrease in the

total energy of the atom, and this lost energy will reappear in the form of heat.

All the compounds of radium are found to be spontaneously luminous. The larger portion of this luminosity is lost in damp air, but it is recovered on drying the salts. It does not appear to be affected by temperature, being as bright at the temperature of liquid air as at ordinary temperatures. Even solutions of radium salts show traces of luminosity. When crystallisation is going on in such a solution, the crystals can be clearly distinguished in the liquid by their greater luminosity.

A remarkable property possessed by radium, thorium, and actinium, but not by uranium or polonium, is that of continuously emitting a material emanation having the properties of a radioactive gas. Like ordinary gases, it is able to diffuse rapidly through gases and porous substances, and it may be separated from other gases with which it may be mixed, by the condensing action of extreme cold. The emanations from these three substances are all of the same general character, but in the case of the radium emanation, the phenomena are much more marked, owing to the relatively great activity of that element.

The emanation can be prevented from escaping from the radioactive substance by covering the latter with a thin plate of mica. It can be carried away by a current of air or other gas, and it can be passed through a plug of cotton wool, or bubbled through solutions, without any loss of radioactivity. Its behaviour is, therefore, very different from that of the ions produced in a gas by the radiation from a radioactive substance, for, under similar conditions, the latter lose their charges completely. In some cases these emanations are continuously emitted from the radioactive substance into the surrounding gases, while in others the emanations become occluded within the material, and can only be released by the action of heat or by dissolving the substance. The emanations are produced in such extremely minute quantities that it has not been found possible to definitely establish their nature by the usual chemical tests. It has, however, been shown that the emanations remain unchanged in amount in passing through various acid solutions, through a platinum tube heated electrically to the highest temperature obtainable, or over a layer of red-hot chromate of lead contained in a glass tube. The emanations have also been passed without change, carried by a current of hydrogen, over red-hot magnesium powder and red-hot palladium-black, and, carried by a current of carbonic acid gas, over a layer of red-hot zinc dust. The only known gases capable of remaining unchanged in amount under all these conditions are the members of the argon family recently discovered by Lord

Rayleigh and Sir William Ramsay as constituents of the atmosphere, where they exist only in minute quantities. In their properties they closely resemble nitrogen, but are even more inert than that gas.

A very important discovery was made by Sir William Ramsay and Mr F. Soddy in 1903, while examining the gases evolved from a small quantity of radium bromide which had been kept for a long time in a solid state. After removing the oxygen and hydrogen, an examination by means of the spectroscope showed the presence of carbonic acid. After freezing out this gas together with the emanation, the residue was found to give, under spectroscopic examination, one of the characteristic lines of the element helium; and on repeating the experiment with a further quantity of the radium salt, the spectrum was found to be that of pure helium. Some further researches were then made with the object of ascertaining whether the helium was really produced by the disintegration of the radium, or whether the appearance of its spectrum was due to traces of helium already existing in association with the radium salt. A quantity of the emanation was first conveyed, with the aid of pure oxygen, into a tube cooled by liquid air, and the oxygen was then pumped out, and the tube removed from the liquid air and examined under the spectroscope. No trace of the helium spectrum could then be detected, but, after standing for four days, some of the characteristic lines of helium appeared. On the fifth day various other lines known to belong to the spectrum of helium made their appearance. These results show conclusively that helium was developed by a spontaneous change in the radium emanation.

Sir William Dewar, as the result of a long series of interesting experiments, communicated to the British Association in 1908, has found that helium is continually produced from radium, at the rate of 0·37 milligramme a day from a gramme of radium.

Further evidences of transmutation were obtained by Ramsay and Cameron in 1907. They found that when the radium emanation was in contact with, and partially dissolved in, water, the inert gas produced by the change consisted mainly of neon, a gas of the argon group, only a trace of helium being observable. When the emanation was left in contact with a solution of the sulphate or nitrate of copper, the main product of the change was found to be argon, with perhaps traces of neon, but no helium. After removing the copper from the solution, the spectroscopic examination of the residue revealed the spectra of both sodium and calcium, together with the red line characteristic of lithium. Portions of the copper solutions which had not been exposed to the emanation showed no traces whatever of lithium. The

authors point out that the gases formed by the disintegration of radium emanation belong to the same group of elements as the emanation itself, and it would appear as though copper also undergoes disintegration, yielding one or more of the lower metals of the copper group.

In connection with these results it should be noted that traces of sodium and lithium are found to be present in practically all substances, and that it is an extremely difficult matter to eliminate them entirely. Madame Curie and Mademoiselle Gleditsch repeated this experiment in 1908, and their results led them to the conclusion that the sodium and lithium observed were probably already present in the apparatus or reagents employed. The radioactivity of these emanations does not remain constant, but gradually decays. Rutherford and Soddy found, in 1903, that the activity of radium emanation fell to half its value in 3·71 days. Ramsay and Soddy, in 1904, found that the volume of radium emanation decreased according to the same law, and that it fell to half its value in about four days, a solid deposit being formed at the same time on the interior surface of the tube.

A very remarkable property of radium, thorium, and actinium is their power of causing temporary radioactivity in neighbouring bodies. It was first observed in 1899 by Monsieur and Madame Curie in the case of radium, and by Rutherford in the case of thorium. In 1903 Debierne found that it was also possessed by actinium. A body which has been exposed for some time to radiation from one of these substances is found to behave as if its surface were covered with a film of radioactive matter, and emits radiation capable of ionising a gas or affecting a photographic plate. The effect gradually disappears on removal of the exciting substance, the time of decay being several days when due to thorium, but only a few hours when due to radium. This induced radioactivity is quite independent of the nature of the substance in which it is excited, and this suggests that the phenomenon is actually due to the deposition of a film of radioactive substance. This view is confirmed by an experiment made by Miss Gates in 1903, in which the induced radioactivity on a fine platinum wire was driven off when the wire was heated to whiteness, and was transferred to surrounding bodies. It has since been conclusively shown that the result was due simply to the volatility of the radioactive deposit. The effect is specially marked in the case ef negatively electrified substances.

The experiments of Wilson, and of Elster and Geitel, on the rate of leakage of an electrical charge in air under varying circumstances, described at the beginning of Chapter IV., led the latter observers to suppose that the results obtained might be due to the

presence of radioactive material. To test this, a wire some 20 metres long was exposed to the air for some hours at a considerable height from the ground, and was kept negatively charged to a high potential by connection with the negative terminal of an influence machine. The wire was then found to largely increase the rate of leakage from an electroscope. No such effect was produced in the case of a positively charged wire. The wire was found to lose its activity after a time, and when rubbed with leather moistened with ammonia, the activity was transferred to the leather. The effect was quite independent of the metal of which the wire was formed, and measurements of the rate of decay showed it to be due to the active deposit from radium emanation, which was thus shown to be present in the atmosphere. In 1902 Sir J. J. Thomson demonstrated the presence of radioactive material in the ordinary Cambridge water supply, and C. T. R. Wilson obtained evidence of its presence in freshly fallen rain and snow, evidently brought down from the atmosphere; in the same year Elster and Geitel showed that the radioactive phenomena observed in caves were due to the presence of radioactive material in the soil. In 1906 they found that a large quantity of radioactive emanation could be obtained by sucking air through clay.

Experiments made in 1902, by Elster and Geitel in Germany, and by Rutherford and Allan in Canada, showed that atmospheric radioactivity was not sensibly affected by moisture, and that it was greatest in a strong wind. The former observers then carried out a detailed examination, extending over a period of twelve months, of the effect of meteorological conditions on the amount of induced radioactivity obtainable from the atmosphere, and they found that the extreme values attained varied in the ratio of 16 to 1. The maximum results were obtained during fog, doubtless owing to the particles of water becoming centres for the deposition of radioactive matter. The activity was greater at lower temperatures and also at low barometric pressures.

Investigations made by numerous observers, some of which are further considered in Chapter XVIII., have now shown that radium and other radioactive substances are very widely distributed in small quantities throughout the earth, and that practically all known terrestrial substances, when tested by sufficiently delicate methods, exhibit some traces of radioactivity. The question whether this universal distribution of radioactivity can be entirely accounted for by the presence of traces of what are known as radioactive substances, or whether the property of radioactivity belongs essentially to every form of matter, is one of the greatest theoretical interest and importance. The view of the constitution of matter which is set forth in this work would lead us to antici-

pate that radioactivity will prove to be, to a greater or less extent, a property of every form of matter, as will be clearly evident after reading Chapter XXI. It should be noted that, with our present means of investigating radioactive phenomena, it would be impossible to obtain definite proof that radioactivity in some shape is not a property of every kind of matter, as we saw in Chapter VII. that there is at present no known method of detecting the presence of alpha and beta particles unless their velocities exceed certain limits. For the same reason we cannot expect, even if radioactivity is a universal property of matter, to be able to obtain definite evidence of its existence in the case of every substance examined. Sufficient evidence has, however, already been obtained to afford good grounds for supposing that the emission of observably radioactive rays is by no means confined to the comparatively small class of substances designated under the name of radioactive bodies.

If air contained in a closed vessel be tested for electrical conductivity, a certain amount of ionisation will invariably be observed. The most obvious cause for this is the radium emanation which is always present, to a certain extent, in the atmosphere. If, however, this were the sole cause, the ionisation should disappear in the course of a few days, owing to the disintegration of the emanation contained within the vessel. This is not found to be the case. On the contrary, when a vessel is filled with pure air, the ionisation is found to increase for some time. Another source of some of the ionisation produced inside the vessel is the active deposit which must be formed on the outer surface from radium emanation present in the external air, as some of the highly penetrative radiation emitted by this deposit must reach the interior of the vessel and give rise to ionisation. The ionisation which could be produced in this way is, however, quite insufficient to give rise to the observed results. Cooke showed in 1903 that even when the air was contained in a metallic vessel surrounded with very thick lead screens, it was impossible to reduce the ionisation to less than about 30 per cent. of its normal amount; and it had previously been shown by him, and independently by Strutt, and M'Lennan and Burton, that the amount of ionisation varied with the material of the containing vessel, even when the interior was carefully cleaned, in order to remove any possible traces of radium deposit, before beginning the experiments.

These results showed conclusively that some of the ionisation was due to radiation emitted by the material of the vessel, and therefore that all the metals experimented with emitted such radiation. We should prove that some of this radiation must be due to the radioactivity of the metal of which the containing vessel is formed

if we could show that its two other possible sources are insufficient to account for the observed ionisation. The first of these is the presence of impurities in the metal of which the vessel is formed of small quantities of radioactive substances. The possibility of the presence of uranium in sufficient quantity is excluded by the fact that it would be revealed by the ordinary methods of chemical analysis, and researches made on a number of metals by Sir J. J. Thomson, Righi, A. Wood, and N. R. Campbell, which failed to reveal a trace of any of the radioactive substances, show that the radiation could not, to any appreciable extent, be due to this source. It is true that Elster and Geitel succeeded, by very refined methods, in demonstrating the presence of traces of polonium in lead, but in quantities so minute as in no way to affect our present problem. The second possibility is that the ionisation may be fully accounted for by secondary radiation emitted by the material of the containing vessel, but deriving its energy from the impact of external radiation upon its outer surface. There is plenty of evidence to show that some of the ionisation produced by metals is due to radiation of this kind, but it is difficult to suppose that it can account for the whole effect, particularly when the containing vessel is surrounded by a considerable thickness of lead, as in Cooke's experiments. Some of Campbell's experiments referred to above afford further evidence in support of the hypothesis that the ionisation is largely due to radioactivity of the metal itself. He found the absorption of the radiation by air to be comparable to that of the alpha rays from radium, and, on the assumption that the radiation was actually of the alpha type, he succeeded in measuring the ranges of the rays emitted by various metals. These ranges were found to have a definite value for each metal, as was to be anticipated if the radiation were due to the radioactivity of the metal itself. His later experiments (*Phil. Mag.*, Feb. 1906) on lead, copper, aluminium, zinc, iron, platinum, tin, silver, and gold, strongly confirm his earlier conclusions, and appear to prove beyond doubt that the emission of ionising radiation is an inherent property of all the metals investigated. The definite value of the intrinsic absorbable radiation from each square centimetre of the surface of the metal, a quantity known as Bragg's constant, for each metal, and its variation with the nature of the metal, shows that the emission cannot be due to the presence of radioactive impurities common to all. A very striking result is that, in every case except aluminium, the penetration of the rays from these metals was found to be greater than that of the most penetrating rays from radium. Secondary, as well as primary, radiation was observed in all cases except aluminium and lead.

Nothing is positively known regarding the origin of the atomic disintegration which we have taken as being the cause of the phenomena of radioactivity, but if we adopt the theory of atomic constitution developed in this work, the matter for wonder is, not that disintegration should occur, but that it does not occur more frequently, for, whenever the velocities of the sub-atoms are undergoing any changes, the atoms must radiate ether waves. Their energy will therefore gradually diminish, and systems of revolving electrons, which are stable when moving with sufficiently high velocities, may be expected to become unstable as the velocities diminish.

A very simple and beautiful method of obtaining a visible demonstration of the luminous effects produced by exposing solids to the bombardment of the alpha particles emitted by radioactive substances was discovered in 1903 by Sir William Crookes, and also, quite independently, by Elster and Geitel. This consists simply in covering any convenient flat surface with a layer of phosphorescent sulphide of zinc, and exposing the screen, so formed, to the impact of the alpha rays, with the result that it becomes brightly illuminated. When the surface of the screen is examined through a magnifying glass, it is seen that the illumination is not continuous, but is due to a great number of distinct scintillating points of light.

Sir William Crookes devised a simple little piece of apparatus which he called the "Spinthariscope," for conveniently exhibiting the phenomenon, and any reader who cares to repeat the experiment for himself can now purchase a spinthariscope at the cost of a few shillings from any dealer in scientific apparatus. A small zinc sulphide screen is fixed into one end of a short brass tube, a small piece of metal which has been dipped into a solution of a radium salt is then fixed at a distance of a few millimetres from it, and the screen is looked at through a lens placed at the other end of the tube. The screen then appears as a dark field covered with brilliant points of light which flash out and disappear with great rapidity. If the distance of the exciting body from the screen is gradually increased, the number of points of light to be seen at any one instant rapidly diminishes, and when the distance is increased to a few centimetres, only an occasional flash will be seen.

Zinc sulphide is extremely sensitive to mechanical shocks, and a momentary luminosity can be produced by drawing the point of a pin or a needle across the screen, or even by allowing a current of air to impinge on it. Experiments made by Becquerel in 1903 led to the conclusion that the scintillations are caused by irregular cleavages in the crystals of sulphide of zinc due to the impact of

the alpha ray particles. Similar scintillations can be produced by crushing a crystal of the sulphide.

The experiment is an extremely beautiful one, and vividly impresses the fact that the observer is witnessing the effects produced by the impact on the screen of single atoms of matter projected with enormous velocities. Rutherford has suggested that every scintillation that occurs when radium emanation is in proximity to a zinc sulphide screen is due to the impact of an atom of helium. He considers that the effect is not simply a mechanical one, but that chemical dissociation accompanies each impact, followed immediately by a recombination, and that it is to this recombination that the flash is due. In a communication to the Dublin Meeting of the British Association in 1908, he described some experiments which he had made with a view of determining the efficiency of the transformation of the energy of the particle into light energy. He measured, by a direct photometric method, the luminosity of a zinc sulphide screen which had been enclosed, together with 200 milligrammes of radium bromide, in a glass tube, with the result that this was between $\frac{1}{200}$ and $\frac{1}{80}$ of a candle-power. This would give about 80 per cent. as the efficiency of transformation of the energy of the alpha particles into light energy.

Many other substances, in addition to zinc sulphide, may be made to fluoresce by the action of the radiations from radium and other radioactive substances, as, for example, the platinocyanides which were referred to in Chapter VII. as being used for covering the luminous fluorescent screens employed in examining objects by the aid of the Röntgen rays. The same property is exhibited to a very marked degree by diamonds, rubies, sapphires, and some other gems, as also by crystals of fluorspar. Certain kinds of glass, and many other substances, also show it, though less markedly. Bodies which fluoresce under the action of the rays frequently undergo other changes, such as changes of colour. It was very soon observed that glass bottles containing salts of radium became gradually darkened, potash glass becoming tinged with brown, while soda glass acquires a beautiful violet colour. The coloration may be removed by strongly heating the glass. Colourless crystals of fluorspar may be coloured by prolonged exposure to radium radiation, and crystals already naturally coloured violet, blue, green or yellow may have the colours intensified or changed in tint in the same manner. It is interesting to note that the naturally coloured crystals, as well as those coloured by exposure to the radioactive rays, can be made colourless by heating to a sufficiently high temperature. A very curious phenomenon has been observed in the case of the

dark violet crystals found in the blue-john mine in Derbyshire. The colour of these crystals is frequently so dark that large crystals appear almost black. When such a crystal is heated until it becomes colourless, and is then subjected to the radium radiation, it assumes a beautiful sky-blue tint, and at the same time it becomes strongly fluorescent, the fluorescence continuing for some days after removal from the influence of the rays.

Mr S. Walter, in a paper on the artificial production of rubies and other gems, in *The Times Engineering Supplement* of the 15th of April 1908, gives an account of some very interesting experiments carried out by E. Ambrecht, in which the latter endeavoured to apply the darkening effect of the radium rays to the manufacture of artificial gems, and a very considerable success has been already achieved in this direction. The raw material employed consists of the colourless crystals of corundum, commonly known as white sapphires, and which are formed of practically pure alumina.

As early as 1857 Gautin succeeded in obtaining rubies by fusing hydroxide of aluminium, previously moistened by bichromate of potash, before the oxyhydrogen blow-pipe. Since that time other workers have obtained like results by a variety of methods, the most noteworthy being, perhaps, those achieved by Fremy and Feil, who fused alumina and lead oxide together in a siliceous crucible at a bright red heat, and obtained in this manner a vitreous mass from which they were subsequently able to extract colourless crystals of corundum, and on the addition of bichromate of potash or oxide of cobalt to the materials, they obtained fine crystals of artificial ruby and sapphire, respectively.

The procedure adopted by Ambrecht is to expose the crystals of corundum, contained in a steel cylinder, to bombardment from the particles thrown off from a small quantity of radium bromide, enclosed within the cylinder, the crystals having been first cut and polished. In the course of a few weeks of such treatment it is found that many, if not all, the crystals have become coloured, some being pale blue, others pale pink, while the majority are usually brownish orange. The stones are then separated into groups according to their different colours, and each group is subjected to the prolonged action of the radium. The coloration then becomes continually more pronounced, until ultimately the crystals become almost indistinguishable from actual gems; rubies, blue sapphires, or oriental topazes, as the case may be.

CHAPTER XVI.

THE THREE PRINCIPAL TYPES OF RADIOACTIVITY.

THE three types of radiation emitted from radioactive substances are of such interest and importance as to make it worth while to consider them in some detail.

According to Rutherford, if the ionising power is measured by the ionisation produced when a thin layer of unscreened radioactive material is spread on the lower of two parallel plates 5 centimetres apart, and the penetrating power is measured by comparing the thickness of aluminium foil required to reduce the radiation of each type to half its initial amount, then the relative ionising and penetrating powers are represented roughly in the following table :—

Radiation.	Relative power of ionisation.	Relative power of penetration.
Alpha rays . . .	10,000	1
Beta rays . . .	100	100
Gamma rays . .	1	10,000

so that the relative power of penetration is approximately in the inverse proportion of the relative ionisation. These numbers must not, however, be considered as representing anything more than very rough approximations, for the penetrating power varies considerably with the nature of the radioactive substance forming the source of the rays.

The Alpha Rays.—The true nature of these rays was not determined until about three years after the discovery, in the latter part of 1899, of the magnetic deviability of the beta rays. We shall see that the alpha rays play a far more important part in radioactive phenomena than the beta rays, and contribute the greater part of the energy emitted in the form of ionising radiation. It was, however, natural that the beta rays should almost

monopolise the attention of experimenters in the earlier stages of research into the phenomena of radioactivity, on account of their much greater penetrative power and capacity of producing phosphorescence in many substances.

Rutherford, in 1903, succeeded in demonstrating the deviability of these rays by magnetic and electric fields, and his measurements showed that the alpha rays consist of positively charged particles projected with about a tenth of the velocity of light. They are, in fact, of a similar nature to the canal rays described in Chapter VII.

The apparatus employed for measuring the magnetic deviability was simple and ingenious. A thin layer of radium, of very high activity, was spread over the bottom of a deep box of rectangular section. The rays from this layer passed upwards between a number of thin vertical brass plates forming a series of equidistant parallel partitions across the box, and extending nearly to the bottom. This was effected by cutting equidistant grooves in a pair of brass plates forming opposite sides of the box, and slipping the thin plates into these grooves. The magnetic field was applied in a direction parallel to the thin plates, which would tend to produce a deflection in a direction perpendicular to them, so that a very small deviation would be sufficient to cause the rays to impinge on the plates and be absorbed. The top of the box was covered by very thin aluminium leaf through which the rays passed upwards into a closed metallic vessel containing a very sensitive negatively charged gold-leaf electroscope, the rate of leakage of which measured the ionisation produced by the rays in the upper vessel. When the magnetic field reached a very high intensity this rate of leakage was found to be diminished, showing that some of the rays were being stopped, which could only be due to their being deflected sufficiently to be driven on to the thin brass plates and absorbed.

In order to prevent any upward diffusion of emanation from the radium into the testing vessel a current of hydrogen was kept flowing into the latter, passing through the porous aluminium leaf, then flowing downwards between the brass plates, and finally passing into the air through an opening at the bottom of the lower vessel. Hydrogen was employed in preference to air, because hydrogen absorbs the alpha rays less than air, and the ionisation due to the beta and gamma rays is much less in hydrogen than in air. The parts of the ionisation due to the beta and gamma rays were separately determined and allowed for by covering the layer of radium with a piece of mica of sufficient thickness to absorb all the alpha rays, but which did not absorb any appreciable proportions of the beta and gamma rays.

The direction of the deviation was determined by closing half

the top of each slit by means of brass strips, all the strips being on the same side of the slits, so that there would be greater absorption when the deviation was towards the covered sides than when towards the open ones, and therefore the amount of the absorption would vary with the direction of the magnetic field. In this way it was found that the deviation was in the opposite direction to that undergone by the cathode rays. As the cathode rays are known to consist of negatively charged particles, this result shows that the alpha rays consist of particles carrying positive charges.

The same apparatus and procedure were employed in determining the deviation of the rays in the electric field, except that the grooved brass plates were replaced by plates of ebonite. Alternate plates were then connected together, and charged to a high potential difference by means of a small battery of accumulators.

Mackenzie, in 1905, and Rutherford and Hahn in 1906, made a long series of investigations of the alpha radiation from different radioactive substances, and found the ratio of the charge to the mass invariably constant. On the assumption that the charge was the same as that of the hydrogen ion, these observers, together with Des Coudres, found that the mass must be very nearly double that of an atom of hydrogen, or the same as that of the molecule of hydrogen. Rutherford pointed out that, if the charge were twice as great, the mass would be about the same as that of the atom of helium, and propounded the theory that the alpha particles were actually atoms of helium carrying such charges. As it had been held previously that atomic charges were in all cases equal to that carried by the hydrogen ion in electrolysis, this theory was not favourably received by physicists, but some recent investigations by Rutherford and Geiger appear to offer such strong confirmation of this theory as almost to amount to a demonstration. Quite recently, but prior to the last-mentioned researches, Bragg, Kleeman, and others have shown that the initial velocities of projection are, within the limits of errors of observation, identical for all alpha particles emitted by the same radioactive substance. For different radioactive substances, the initial velocities vary between 10,000, and 14,000 miles a second. Rutherford and Geiger's investigations, referred to above, were communicated to the Royal Society in June 1908, and the object in view was to endeavour actually to count the alpha particles emitted from a given mass of some radioactive substance in a known time. As the total electrification caused by these particles could be easily measured, the question whether the alpha particle carried a single or a double charge could then be determined by simply dividing the total electrification by the number of particles.

If it could be assumed that one, and only one, scintillation is produced by the impact of each alpha particle upon a zinc sulphide screen, the counting might be effected in this manner, as it is easy to arrange an experiment so that the scintillations may be counted by the aid of a microscope. Some observations made in this way were found to be confirmatory of the hypothesis that each particle carried a double charge, on the assumption that every alpha particle gave rise to one, and only one, scintillation. The validity of this assumption appeared, however, too doubtful for the conclusion to be accepted without independent evidence obtained in some other way. After overcoming various experimental difficulties, Rutherford and Geiger succeeded in doing this. An arrangement was devised by means of which alpha particles could be fired through a gas at low pressure exposed to an electric force as high as could be employed without sparking taking place. The result of this electric force was to increase the ionisation several thousand times. Consequently, the sudden current through the gas due to the entrance of an alpha particle through a thin sheet of mica into the testing vessel was so much increased as to give rise to an easily observable throw of the needle of an ordinary electrometer. The results obtained were in close agreement with those derived from the scintillation method, indicating that either method of counting might be employed with confidence.

Rutherford and Royds (*Phil. Mag.*, Feb. 1909) have obtained further confirmation of the identity of alpha particles with helium by means of the steady development of the spectrum of helium in a vessel continuously fired into by alpha particles. The particles were emitted through the walls of a fine glass tube containing radium emanation.

The kinetic energy of the alpha particles is enormous compared with their masses. Its value for a group of particles would be about six hundred million times as great as that of a rifle bullet of the same mass as the aggregate mass of the group, and travelling with a velocity of a thousand metres a second. This must be regarded as the principal source of the heating effects referred to in the preceding chapter as produced by radium.

The ionisation produced by alpha particles, after traversing in succession a number of extremely thin sheets of aluminium foil of equal thickness, was shown by Rutherford and Miss Brooks in 1902 to be very nearly halved by each additional thickness until it was reduced to about 6 per cent. of its initial value, after which it began to diminish much more rapidly and soon fell to zero. When the aluminium was replaced by other substances, solid, liquid, or gaseous, experiments made by various observers led to

the conclusion that, with some marked exceptions, this apparent absorption of the alpha particles by various kinds of matter was generally proportional to the density of the substance as well as to its thickness, *i.e.* to the quantity of matter traversed by the particles. The beta rays were practically eliminated temporarily from the radium chloride employed in these experiments, by solution of the salt in water and subsequent evaporation.

The most obvious means of accounting for this apparent absorption lay in the retardation of the particles which must necessarily accompany the loss of energy expended in collisions with the molecules of the matter traversed, which collisions are the cause of the ionisation in the case of gaseous media. A long and careful series of researches by Bragg and Kleeman in 1904 and 1905 resulted in explaining the apparent anomalies which had prevented the earlier observers from accepting this solution, and had led to the suggestion of various tentative hypotheses. These anomalies were shown by them to be due to the fact that radium emits no less than four distinct sets of alpha particles with distinct ranges consequent on differing initial velocities of projection. These are due to the formation of distinct radioactive products, and the key to the solution of the difficulty introduced by their presence was found in the fact that they could be temporarily eliminated by heating to redness the radium salt employed. Similar phenomena are exhibited by uranium, thorium, and actinium. After this elimination of all except one set of alpha particles, the ionising effect in air was found to increase at first, and then rapidly diminish, as the distance from the radium layer increased. This was in accordance with the conclusions arrived at by Townsend in 1901, and confirmed by Durach a couple of years later, that the ionising effect of a moving particle attains a maximum value for a certain critical velocity, and falls off as this is either increased or diminished. It was shown by Rutherford in 1905 that when the velocity of the single set of alpha rays from radium was reduced to below 60 per cent. of its initial value it no longer gave rise either to ionisation or to phosphorescent or photographic effects. We should therefore have no means of detecting the presence of particles moving with less than this speed.

Rutherford (*Phil. Mag.*, Jan. 1907) finds that the velocity of expulsion of alpha particles from various radioactive substances all lie within a fairly narrow range, viz. between $1{\cdot}56 \times 10^9$ and $2{\cdot}25 \times 10^9$ centimetres per second. He has also shown that the particles cease to give rise to ionisation when their velocity falls below a value of about $\cdot 82 \times 10^9$ centimetres per second, so that the velocity of expulsion varies generally between two and three

times this critical velocity. As a particle escaping below this velocity would be difficult to detect, he remarks that, if the speed of projection had been only half as great, on the average, as the observed values, they might have escaped detection altogether.

It is found, as a general rule, that the velocity of expulsion increases progressively as the period of transformation decreases, which would appear to indicate that the velocity of expulsion is greatest for the least stable substances. There are some noteworthy apparent exceptions to this general law, but Rutherford considers that they are most probably to be attributed to the existence of two successive changes which have not yet been discriminated from each other—a rayless change of comparatively long period, and a ray change of very short period. Rutherford finds the energy of the alpha particles emitted by radium compounds to have an average value of about $5{\cdot}9 \times 10^{-6}$ erg, that of the beta particles averaging about 7×10^{-8} erg.

The Beta Rays.—The magnetic deviability of the beta rays was discovered independently, and almost simultaneously, in the year 1899 by Giesel, Meyer and von Schweidler, and Becquerel. The latter observer used radium, and the others mixtures of polonium and radium. Becquerel proceeded to a detailed investigation of the deviability of the rays by magnetic and electric fields. A photographic plate, enveloped in black paper to prevent its being affected by light, was exposed to the action of a radium salt contained in a leaden vessel. In some of the experiments a mica screen was placed in the path of the rays, and gave rise to a sharply defined shadow on the plate, while in others the rays were made to pass through a narrow slit, of which a well-defined image was formed on the plate. When a strong magnetic field was produced, with its lines of force parallel to the plate and to the length of the slit or mica strip casting the shadow, or a strong electric field, with its lines of force parallel to the plate and perpendicular to the length of the slit or mica strip, the shadow or image of the slit, respectively, was deflected to one side or the other according to the direction of the field, the direction of displacement being such as to show that the rays consisted of negatively charged particles. The deflected image of the mica plate or slit was no longer sharply defined, but appeared blurred at the edges, showing that the charged particles were deflected by varying amounts, indicating a considerable variation in their velocities. It was found that the particles which were most deflected, and which were therefore moving most slowly, were the first to undergo apparent absorption in passing through successive thicknesses of matter. Careful measurements led to the conclusion that the ratio of the electric charge on a particle to its mass had an

average value very nearly the same as in the case of the cathode ray particles, but that the velocities of projection were much greater, varying from about half to more than nine-tenths of the velocity of light.

The experimental method employed by Kaufmann in investigating the velocity of the beta particles was to expose a photographic plate wrapped in thin aluminium foil to the action of the narrow pencil of rays passing through a tube about a fifth of a millimetre in diameter inserted in the top of a metal box containing a salt of radium. The apparatus was placed in an exhausted vessel, and subjected simultaneously to strong magnetic and electric fields with their lines of force parallel to each other and to the photographic plate. The result was to produce magnetic and electric displacements of the particles, both displacements being parallel to the photographic plate but at right angles to each other, giving rise to a curved line on the plate. This experimentally obtained curve was then compared with the curve calculated from electromagnetic theory.

The observed and calculated values of the ratio of the charge of an electron to its mass at various velocities extending up to over 96 per cent. of the velocity of light were found to be in very close agreement, the differences averaging a little over 1 per cent., some of the observed results being slightly greater and some slightly less than the calculated ones. The actual differences varied from nothing to 7·8 per cent. The following table, in which v is the ratio of the velocity of the electron to the velocity of light, and m the corresponding ratio of the actual mass of the moving electron to its mass at rest, gives the calculated value of m for various values of v.

v	Less than ·1	·1	·5	·9	·99	·999	·9999	·999999
m	1·0	1·015	1·22	1·81	3·28	4·96	6·68	10·1

It will be noted that, although the mass becomes infinite when the velocity of the electron attains the velocity of light, yet when the velocity of the electron is only less than that of light by one part in a million, the mass only reaches about ten times its value when at rest.

The absorption of the beta rays by matter is, as in the case of the alpha rays, generally proportional to the quantity of matter traversed. Lead is a noteworthy exception, having an absorbing power nearly twice as great, in proportion to its density, as would be anticipated from the general law. The absorptive power of tin,

in proportion to its density, is even more abnormally high for beta rays, although it is abnormally low for alpha rays.

The Gamma Rays.—The emission of these very penetrating rays by salts of radium was observed by Villard in 1900 by means of their photographic effects. They also produce ionisation in gases, and this provides the most convenient means for their investigation. The alpha and beta rays from radium are completely absorbed by a sheet of lead a centimetre in thickness, but, according to Rutherford, the presence of gamma rays from 30 milligrammes of radium bromide can be observed by means of an electroscope after passing through 30 centimetres of iron, which would be equivalent to nearly 11 centimetres of lead. Their absorption by matter of various kinds is governed by the same general law as in the case of the alpha and beta rays, with similar well-marked exceptions.

They show many resemblances to the hardest Röntgen rays, and they are generally supposed to consist, like them, of ether pulses, and it is to be expected that such pulses would be set up at the sudden liberation of electrons, as well as at their sudden stoppage. The latter is accepted as the origin of the Röntgen rays, and the much greater penetrative power, or *hardness*, of the gamma rays as compared with these would only be in accordance with what would be anticipated, assuming this to be their nature, from the high velocities of projection of the alpha and beta particles as compared with those forming the canal rays and cathode rays respectively. They certainly do not, like the alpha and beta rays, carry electric charges. Some experiments made by Paschen in 1904 appeared at first to indicate that such charges were carried by these rays, but Eve showed conclusively in the same year that the phenomena which appeared to suggest such a conclusion were really due to secondary beta rays starting from the points of impact of the gamma rays upon solid obstacles.

W. H. Bragg suggested in 1907 that the gamma rays may consist of uncharged particles or doublets formed by the coalescence of an alpha particle with a beta particle, and succeeded in showing that many of the properties of the rays could be accounted for on this hypothesis. His conclusions have, however, been contested by G. H. Barkla, and the resulting discussion appears so far rather to favour the older ether-pulse theory, but the matter cannot as yet be considered as determined.

CHAPTER XVII.

TRANSMUTATIONS OF RADIOACTIVE SUBSTANCES.

Uranium.—Reference was made in Chapter XV. to the discovery that uranium consists of two distinct constituents. This was made by Sir William Crookes in 1900. Uranium can be precipitated from its solutions by the addition of a solution of ammonium carbonate, which must be added slowly and carefully, as the precipitate is redissolved when the carbonate is added in excess. Crookes found that a small residue was left undissolved after this treatment. This residue was of a light brown colour, and many hundred times more radioactive, when tested photographically, than uranium itself. Crookes called it uranium X. Experiments of a very similar character were made about the same time by Becquerel, who succeeded in so completely freeing the uranium from uranium X that the former failed to show any effect on a photographic plate covered with black paper in order to prevent it being acted on by light. In these experiments a solution of barium chloride was added to one of sulphate of uranium, with the result of precipitating the barium in the form of barium sulphate. This latter was found to be strongly radioactive, while the radioactivity of the uranium was greatly diminished. This process was repeated a number of times with the original uranium solution until the salt, when dried, failed to affect the photographic plate. The photographically inactive uranium and the active barium were then laid aside for about a year, and again examined, when it was found that the activity of the barium had entirely disappeared, while that of the uranium had been completely recovered. When uranium and uranium X were examined electrically by their power of producing ionisation it was found that the activity of uranium X was comparatively small, and that its removal caused very little diminution in the activity of the uranium. These results were fully explained in 1902 by researches carried out by Soddy and by Rutherford and Grier, from which it appeared that uranium, when freed from uranium X, only emits alpha rays, which are comparatively inactive photographically, but have a

high ionising power, while uranium X emits only beta and gamma rays, which have comparatively feeble ionising powers but a strong photographic action. The conclusion indicated by the earlier results that uranium X is continuously produced by uranium was confirmed, and its rate of decay was shown to be such that its radioactivity fell to half its value in twenty-two days. The production of uranium X from uranium is therefore in every way analogous to that of radium emanation from radium, which was referred to in Chapter XV., except that at ordinary temperatures the former is solid and the latter gaseous. There is at present no direct evidence as to the nature of the product resulting from the breaking up of uranium X, as, in common with most of these transmutations, the quantity obtainable is too small to be detected by means of any of the ordinary chemical tests. There is, however, as we shall see presently, a considerable amount of indirect evidence leading to the conclusion that it consists, at any rate in part, of a substance which is found in nature in association with actinium, and to which the name of ionium has been given. Similar evidence points to the conclusion that ionium becomes further transmuted into radium and its series of emanations, which we shall now go on to consider.

Radium.—It was pointed out in Chapter XVI. that, in the emanation known to be produced from radium, investigation of the alpha rays had indicated the existence of other radioactive bodies. A study of the rates of decay of radium emanation led Rutherford to similar conclusions, and an analysis of the curves representing the observed variations in the rate of decay of the active deposits with the time, measured in various ways and under varying conditions, led to the conclusion that these three different radioactive substances represented successive stages in the transmutation of radium.[1] They were called radium A, radium B, and radium C respectively.

The volatility of the deposit discovered by Miss Gates suggested the possibility of separating these substances by subjecting them to different degrees of temperature. Curie and Danne succeeded in 1904 in separating radium B and radium C in this manner, the former being completely removed from an active wire at a temperature of a little over 600° C., while the latter was not completely dissipated even when the temperature was raised as high as 1300° C.

Madame Curie found in 1903 that, although the greater part of the active deposit disappeared within a few hours after its removal from the emanation from which it had been evolved, there always

[1] See Foster and Porter's *Electricity and Magnetism*, p. 598.

remained a small residual activity, representing not more than $\frac{1}{20,000}$ of the activity of the original deposit, which did not appear to decay with time. This residual activity was investigated by Rutherford, resulting in the discovery of three more radioactive products, which were designated radium D, E, and F respectively. A platinum plate was exposed to a large quantity of radium emanation for a long period, and it was found that, after the lapse of a sufficient time for the complete disappearance of radium C, the residual activity, as measured by the emission of alpha rays, continued to increase throughout a period of eighteen months over which the observations extended. At the end of that time it appeared to be approaching a limit. The activity, when determined by the emission of beta rays, attained its maximum value in about forty days, after which it began to decrease. This suggested that the two types of radiation were due to two distinct products arising from a parent substance which, like radium B, emitted little or no observable radiation. If this were the case it was evident that the rate of decay of the parent product must be extremely slow in comparison with those of the previously known products of radium disintegration. Rutherford found that the alpha ray activity disappeared almost entirely at a temperature of a little over 1000° C., and at still higher temperatures the product emitting beta rays gave indications of being slightly volatile. He also succeeded in separating the product emitting alpha rays from a solution of the active deposit in dilute sulphuric acid.

The rate of decay of radium E, emitting beta rays, was measured, and it was found to be half disintegrated in six days; while radium F, emitting alpha rays, was half disintegrated in 143 days. This latter rate of decay is the same as that of polonium, which also emits only alpha rays. Moreover, polonium was obtained in 1902 by Marckwald, as a black radioactive deposit upon a rod of bismuth dipped in the bismuth chloride solution obtained from uranium residues; believing it to be a new substance, he called it radio-tellurium. Radium F was separated by Rutherford in a similar manner, and it is therefore now generally accepted as being identical with polonium. The time estimated by Rutherford, from an analysis of the curves, for radium D to be half disintegrated is forty years. It was shown by Meyer and Schweidler in 1906 that radium E itself consists of two successively formed products, radium E_1, which emits no rays and is half disintegrated in a period of a little over six days, and radium E_2, which emits beta rays and is half disintegrated in 4·8 days.

Sir William Ramsay informs me that the time taken by radium to fall by disintegration to half its original amount has now

been definitely calculated from the volume of the emanation experimentally determined by R. W. Gray and himself,[1] and found to be 1258 years.

The accompanying table gives a list of the radioactive substances which have so far been shown to be related by direct descent one from the other, with their rates of decay and the rays emitted by each. Each substance in the table is formed by the disintegration of the one immediately preceding it:—

Radioactive substance.	Time of decay to half amount.	Rays emitted.
Uranium ↓	About 10^9 years	Alpha rays.
Uranium X ↓	22 days	Beta and gamma rays.
Ionium ↓	About 25 years	Alpha rays.
Radium ↓	1258 years	Alpha rays.
Radium emanation ↓	3·7 to 4 days	Alpha rays.
Radium A ↓	3 minutes	Alpha rays.
Radium B ↓	26 minutes	Slow beta rays.
Radium C ↓	19 minutes	Alpha, beta, and gamma rays.
Radium D ↓	40 years	Rayless.
Radium E_1 ↓	6 days	Rayless.
Radium E_2 ↓	4·8 days	Beta rays.
Polonium	.143 days	Alpha rays.

Thorium and actinium give rise to disintegration products of a general character resembling those of radium, and for this reason and from their association in nature with the radioactive substances now known to be produced by the disintegration of uranium, the existence of some relation between them appears probable. No direct transmutations have, however, as yet been established between thorium and actinium and their products, and the products of uranium.

Thorium.—Rutherford and Soddy found in 1902 that a substance closely analogous to uranium X could be separated

[1] R. W. Gray and Sir William Ramsay, *Chem. Soc. Jour.*, June 1909, p. 1073.

from thorium, and they accordingly called it thorium X. Otto Hahns, working with Sir William Ramsay, found an intermediate stage between these two, consisting of a very slowly decaying substance emitting alpha rays, to which the name of radio-thorium has been given. Thorium X disintegrates into a thorium emanation very much resembling radium emanation in its general characteristics. This in its turn gives rise to an active deposit consisting of two successively formed substances known as thorium A and thorium B. The periods of decay and rays emitted by these substances are given in the accompanying table:—

Radioactive substance.	Time of decay to half amount.	Rays emitted.
Thorium	About 2×10^9 years	Rayless.
↓ Radiothorium	109 years	Alpha rays.
↓ Thorium X	4 days	Alpha rays.
↓ Thorium emanation	53 seconds	Alpha rays.
↓ Thorium A	11 hours	Rayless.
↓ Thorium B	55 minutes	Alpha, beta, and gamma rays.

Actinium.—The products formed by the disintegration show a remarkably close resemblance to those of thorium, as is shown by the accompanying table:—

Radioactive substance.	Time of decay to half amount.	Rays emitted.
Actinium	At least 10^5 years[1]	Rayless.
↓ Radioactinium	19·5 days	Alpha rays.
↓ Actinium X	10·2 days	Alpha rays.
↓ Actinium emanation	3·9 seconds	Alpha rays.
↓ Actinium A	36 minutes	Rayless.
↓ Actinium B	2·15 minutes	Alpha, beta, and gamma rays.

[1] This is the estimate made by B. Keetman (*Jahrbuch der Radioactivität und Elektronik.*, vol. vi., 1909, p. 265) from the results of unsuccessful attempts to obtain actinium as a disintegration product of ionium.

Since radium is continuously undergoing disintegration at such a rate that the amount existing in a certain quantity of mineral at any given time will diminish to half its initial value in a period of twelve hundred and fifty-eight years, it follows that the radium now present on the earth must have been formed very much more recently than the minerals in which it is found. Their age is measured by millions of years, and it is impossible to avoid the conclusion that they would have long since ceased to contain appreciable quantities of radium unless there had been a further formation of radium within the substance of the mineral within comparatively recent times. Geological data practically exclude the assumption that this can have taken place suddenly and then ceased. We are therefore driven to the conclusion that radium is being continuously formed by the disintegration of some other constituent of the mineral. Now, uranium ores are the only known sources of radium in any considerable quantity, and therefore it was suggested by Rutherford and Soddy in 1903 that uranium, which is itself a radioactive substance, might very possibly be the source from which the supply of radium is maintained.

Evidence confirming the truth of this suggestion was obtained independently by Whetham and Soddy in 1904, and was confirmed by Boltwood in 1905. These observers found that radium gradually made its appearance in a solution of a uranium salt, originally free from radium, after it had been kept for some time in a closed vessel free from any possibility of the introduction of external radioactive material, but that the amount so produced was considerably less than could be accounted for on the assumption that it was obtained directly from the disintegration of uranium X. It was suggested that actinium, which is always found to be present in uranium ores, might be the intermediate product between uranium X and radium. The investigation of this was undertaken in 1906 by Boltwood, who had been engaged in following up another line of research, which, as we shall see presently, afforded further confirmation of the hypothesis that radium was produced by the disintegration of uranium. In connection with these researches Boltwood had devised improved methods of analysis of uranium ores, which enabled him to effect a complete separation of the actinium; and also a more accurate method of ascertaining the amount of radium present, by measuring the total amount of radium emanation obtained by dissolving the mineral and boiling the solution. Owing to the fact that nearly the whole of the emanation formed within a solid body remains occluded within its pores, the quantity of emanation obtained in this way is, to a very close degree of approximation, proportional to the amount of radium present.

Boltwood's investigations led him at first to the conclusion that actinium was, as had been suggested, the intermediate product between uranium X and radium, but it was shown by Rutherford in 1907 that this intermediate product was not the actinium itself, but a substance which was precipitated with it, and which has received the name of ionium. Rutherford's result has since been confirmed by Boltwood's further investigations.

In the meantime, the results obtained by investigators following the other line of research referred to above led to the same conclusion that uranium is the source from which the supply of radium is being maintained. M'Coy, whose investigations were published in 1904, analysed a large number of minerals containing uranium, and compared their total activities, with the result that the total activity was always found to be approximately proportional to the quantity of uranium, as determined by the chemical analysis, except in the case of minerals containing an appreciable amount of thorium. This was the result to be anticipated on the hypothesis that the radium and its products present in the minerals were due to the disintegration of uranium. To make this point perfectly clear, let us consider what would occur in the case of a quantity of uranium which had been freed from uranium X and further disintegration products. The amount of uranium X would at first gradually increase until it attained a maximum value when the permanent state was reached in which the amount of uranium X formed in a given time became equal to the amount destroyed by disintegration in the same time. The amount of uranium X would then bear a definite ratio to the quantity of uranium, and the same would hold good with regard to each of the further disintegration products, after the lapse of a sufficient time for each of them to attain its maximum. In this permanent state the total emission of rays of each kind would bear a definite ratio to the quantity of uranium existing unchanged at the time of observation. The activities of different minerals containing no radioactive substances, other than those derived from the uranium, would therefore always bear the same ratio to the quantity of uranium present, provided the activity were measured by means of the same class of radiation in each case. These results were fully confirmed by an extensive series of researches by Boltwood, and also by Strutt and other observers.

In addition to thorium, bismuth, barium, and lead occur constantly in association with uranium, and it has been suggested that some or all of them may be among the disintegration products of uranium. Recent researches by Boltwood, however, show that the three former metals are not found in the proportions which would be required to justify such a conclusion, but he finds that

lead does so to a quite remarkable degree. The comparison in all such cases must be confined to minerals found in the primary rocks, *i.e.* those which have not undergone any change by the action of water subsequent to their formation, as in that case there would be changes in the relative proportions of the various constituents, owing to some of them being dissolved and removed by the percolating water. With regard to such stable substances as bismuth, barium, and lead, a fixed ratio is not to be expected in minerals of widely differing ages, for if any disintegration of these substances is taking place it is only to an imperceptible amount.

Boltwood finds that lead does occur in the requisite fixed proportion to uranium in primary minerals estimated to be of about the same age, and that the proportion of lead relatively to uranium is greater in the case of minerals known to be of comparatively early formation. Boltwood infers from these results that lead is to be regarded as the final disintegration product of uranium.

When we come to consider, in the next chapter, the probable origin of the radioactive substances, we shall see that there is good reason for supposing that there exist in the sun, at the present time, radioactive substances of a higher order, and capable of liberating, by their disintegration, far larger quantities of energy than any now existing on the earth's surface. Some substance of this kind may very possibly have been the common ancestor of both uranium and thorium, which, according to Boltwood's researches, cannot be related to each other in a direct line of descent. The question whether actinium and thorium are related to each other in this manner, or whether actinium and its products form a third line of descent parallel with those of uranium and thorium, is one which it will be of interest to determine.

CHAPTER XVIII.

THE AGES OF THE EARTH AND SUN, AND THE PROBABLE ORIGIN OF RADIOACTIVE SUBSTANCES.

THE phenomena of radioactivity have an important bearing on some cosmic problems of the highest interest, of which the two most important are the age of the earth and that of the sun.

Age of the Earth.—Many years ago Lord Kelvin came to the conclusion that the earth, in the course of cooling down from its original gaseous condition, and after passing through the completely fluid state, must have begun to form a solid crust somewhere between 100,000,000 and 200,000,000 years ago. The geologists, on the other hand, maintained the absolute impossibility of explaining the known facts of geology without assuming that the solid earth had been in existence for a very much longer period. Lord Kelvin's calculations were based upon the application of the wonderful mathematical methods developed by Fourier more than a century ago, primarily for the elucidation of problems in the conduction of heat.

The simplest way to give the non-mathematical reader some conception of the manner in which this problem was attacked will be to consider it first in a simplified form. Imagine a large solid metal sphere having a small hole bored from some point on the surface to the centre, and let this sphere be heated in a furnace until it has attained a practically uniform temperature throughout. Then let the sphere be withdrawn from the furnace and suspended in a place of uniform temperature, and free from air currents. The sphere will then begin to cool in such a way that, at any one time, the temperature at every point on the surface will be the same. To further fix our ideas let us suppose that the sphere is of iron and six feet in diameter, and assume that the first observations are made upon it after a period of ten hours has elapsed since removing it from the furnace. Assume that these observations consist in first taking the surface temperature by means of a thermometer, and then introducing the thermometer into the hole bored to the centre, and taking the tempera-

tures at intervals of, say, every inch from the surface to the centre. It will be found that the temperature at the centre is much higher than that of the surface, and that the temperature of any point within the sphere is greater the nearer that point is to the centre. The application of Fourier's methods to these observations will enable us to write down a differential equation which will determine the relation between the temperature at any point within the sphere and the distance of that point from the centre of the sphere at any given time after the time at which the observations were made. Suppose we now try to solve the equation so obtained, but using a negative value instead of a positive value of the time. This, if successful, would give the temperature at the point selected at that interval of time before, instead of after, the time of observation. As we have assumed that the observations were made ten hours after withdrawing the sphere from the furnace, we should be successful in obtaining a solution for any negative value of the time less than ten hours; but for a greater negative value, say eleven hours, for example, the mathematical expression obtained for the temperature of the given point would become incomprehensible; in other words, the expression would be nonsensical. If we did not know that ten hours was the period which had elapsed between the removal from the furnace and the taking of the observations, the least negative value of the time for which the expression became nonsenical would give us the time of removal. It is not difficult to understand the reason for this. The mathematical relation was obtained on the supposition that the only change taking place in the sphere was that due to the process of cooling. The change in the conditions introduced by removing the sphere from the furnace was not allowed for, and therefore our equation was unable to give us any information on points entirely outside its scope.

The earth is by no means of uniform material throughout, and it is, of course, impossible to make observations extending from the surface to the centre. In fact, the measurements corresponding to those made in the hole bored to the centre of the sphere had to be made in a coal mine whose depth was not much more than a mile, a very small proportion of the four thousand miles or so which would have to be explored if the observations were to be carried out as completely as those made on the metal. All that was therefore possible in the case of the earth was to obtain probable values for the maximum and minimum times respectively. In obtaining the former solution all doubtful quantities would have the extreme possible values given to them in such a direction as would increase the time. The minimum time would then be

obtained by giving to all doubtful quantities extreme values in the direction which would tend to make the time smaller.

In one case the mathematical expression became nonsensical when the time exceeded about 200,000,000 years, in the other when it exceeded about 100,000,000 years. This meant that at some time within these limits something had happened which had introduced a change in the conditions of cooling. Now we know that such a change would be introduced by the passage of the earth from the condition of a liquid to that of a solid, and there was no reason to ascribe the indicated change to any other cause. In these calculations it was, of course, assumed that no heat was being generated in the earth's interior. If such were the case it is quite clear that the conditions of the problem would be entirely altered, so that the solution obtained would no longer be valid.

Now, we have seen that radioactive materials are widely distributed over the earth's surface, and it can be shown that the existence of radium in the earth to the extent of 4·6 parts by weight in a hundred billions would be sufficient to compensate for the earth's present rate of loss of heat by conduction from the interior to the surface, and thence by radiation into space. Some of Elster and Geitel's results afford good ground for believing that the radioactivity of the earth is of the order indicated, and further confirmation of this has been obtained in recent investigations by R. J. Strutt, of which he gave an account in the Friday evening lecture at the Royal Institution on the 27th of March 1908. He pointed out that the mineral pitchblende, the source of radium and other radioactive materials, has, as far as England is concerned, only been found in Cornwall, in veins in the granite and slate. He considers that it is derived from the surrounding granite, as examinations of a very refined character have shown that the granite contains radium to the extent of about one part in a billion. He observes that, minute though this portion is, the total quantity of radium contained at this rate in the outer crust of the earth, to a depth of forty or fifty miles, would be more than sufficient to account for the earth's heat radiation.

It is therefore extremely probable that the present rate at which the earth is losing heat may have continued without any change for very long periods of time, owing to the supply of heat from radioactive matter in the interior of the earth. It is therefore quite possible that the earth may have remained for a very long period of time at a temperature not very different from that of to-day, and therefore that the period during which the temperature has been such as to permit of the existence of

animal and vegetable life may be very much greater than Lord Kelvin's estimate.

Strutt called attention, in the course of his lecture, to the great significance of helium in the investigation of terrestrial radioactivity. We have already seen that helium is produced by the disintegration of radium and other radioactive substances; and, being a stable substance, not subject, as far as our present knowledge extends, to further disintegration, we should expect to be able to detect its presence in any minerals containing radium.

The presence of radium in granite, therefore, led Strutt to search for helium, which he was successful in finding. He also found it in a number of other minerals, such as quartz, for example, the radioactivity of which is not very noteworthy. The helium was usually accompanied by a sufficient quantity of radium to account for its formation. There was, however, one abnormal case, that of beryl, the mineral of which emeralds are composed. Only the smallest traces of radioactivity were found in beryl, and yet it contained large quantities of helium, no less than 17,000 cubic millimetres of helium being obtained from a kilogramme of beryl. It has been suggested as a possible explanation that this mineral might be sufficiently aged for practically the whole of the radium in it to have become disintegrated, leaving only the helium. There do not, however, appear to be any grounds for assuming that beryls were formed earlier than other crystals found in rocks and cavities. Strutt suggests that an explanation might possibly be deduced from some observations recently made by Professor Joly of Dublin. In the course of microscopic examination of thin sections of granite, Joly obtained a section in which a crystal of white mica was very conspicuous, and within the mica was a small crystal of zircon, which is one of the minor constituents of granite. The zircon was surrounded by a dark disc which was very nearly a perfect circle.

There can be little doubt that the darkening of the mineral was due to the alpha rays from radium emanating from the zircon, which is known to be a source of radium, and that the circle was simply a section of the sphere, with a radius equal to the "range" of these alpha rays, circumscribing the portion of the mica within which the rays were effective in giving rise to coloration. This darkening of the mica had been accomplished in the course of geological ages, but Joly succeeded, by the use of concentrated radium, in obtaining a similar effect in his laboratory in the course of a few months.

Strutt's suggestion was that, after the particles had become

incapable of producing coloration, either by diminution in their velocity, or possibly, as Sir J. J. Thomson suggested, by the loss of their electric charges, they might still be able to give rise to the formation of helium. A still simpler explanation is afforded by the definite proof since obtained, as we saw in Chapter XVI., that the alpha particles themselves consist entirely or partially of helium. Strutt then went on to suggest that, since helium has been produced by the gradual disintegration of radioactive substances, and since the rate of production of helium from radium has been determined from laboratory experiments, it appears extremely probable that determinations of the relative quantities of helium existing in various rocks may furnish us with important data regarding the ages of the strata in which they occur.

The Age of the Sun.—It is possible, by measurements of the heat radiated from the sun, to obtain more or less approximate estimates of the total energy emitted from the sun into space during a given period of time.[1] We know from geological evidence alone that the sun must have been emitting heat at a rate certainly not less than at present, and probably greater, for an enormous period of time. According to the older physical theories, all, or practically all, this energy must have been derived from the work done in the gradual contraction of the sun's bulk. Lord Kelvin made some calculations of the amount of energy which would be transformed from the potential to the kinetic form in this manner, and be dissipated in the form of heat during the concentration of the sun from a condition of infinite dispersion. The data for such a problem as this are necessarily somewhat indeterminate, but the conclusions arrived at by Lord Kelvin were that we could be almost certain that the sun has not illuminated the earth for as much as 500,000,000 years, and probably for not more than 100,000,000 years. The geologists demand a very much longer period than this in order to explain the succession of stratified deposits containing the fossilised remains of so many forms of vegetable and animal life. Again, with regard to the future, Lord Kelvin's conclusion was that, unless sources of energy unknown to us at the time of his investigation were in existence, the inhabitants of the earth could not continue to enjoy the light and heat essential to their life for many million years longer.

The discovery of the phenomena of radioactivity has introduced an entirely fresh factor into this problem. Mr W. E. Wilson showed in 1903 that, on the basis of Curie and Laborde's estimate

[1] See Poynting and Thomson's *Heat*, p. 251.

of the heat emission of radium, referred to earlier in this chapter, the presence of 3·6 grammes of radium in each cubic metre of the sun's mass would be sufficient to account for its present rate of emission of energy. From the known average density of the sun it follows that this is equivalent to 2½ parts by weight in a million. No traces of radium have yet been observed in the spectrum of the sun, but the existence of helium was discovered in the sun before it was found in the earth, and strongly suggests the existence of radioactive matter in the former. This suggestion is supported by the information as to the constitution of the sun, obtained from spectroscopic and ordinary telescopic observations which have been carried on for many years, especially during the short periods of total eclipse by the moon.

It should be noted in the first place that the luminous, and therefore visible, stars may be roughly classified, according to the colour of light which they emit, into white stars, such as Sirius, yellow stars, of which the sun is an example, and red stars, such as Betelgeuse. Of these, speaking generally, the white stars are the hottest, and the red stars are the coolest. It is necessary to qualify the general statement as I have done, for we learn from observations on newly formed stars that a star may be surrounded by clouds of small particles, or cosmic dust, as it is called, which may stop a larger proportion of the blue rays than of the red ones, and may therefore cause a star which is of sufficiently high temperature to emit white light, to appear yellow, or even red. Wien's law enables us to obtain an estimate of the temperature by measuring the wave-length of that part of the spectrum of the hot body in which the emission of heat is greatest. The application of this law to the sun would give a temperature of about 5000° C. as that of its visible surface. This result must obviously be too low, as a considerable proportion of the radiation is absorbed by the earth's atmosphere and by that of the sun. When allowance is made for this, the mean temperature of the sun's surface is found to be between 6000° C. and 7000° C., with a most probable value of about 6200° C.

According to the results of observations by Harkányi, it appears probable that the temperature of the white stars Vega and Sirius is about 1000° C. higher than that of the sun, and that that of the red star Betelgeuse is about 2500° C. lower than the sun's temperature. The constituents of the white stars as observed in the spectroscope appear to be mainly hydrogen and helium, with a smaller proportion of oxygen, and very little trace of metals. Metallic spectra predominate in the case of the yellow stars, and are accompanied by those of a few chemical compounds. In

addition to the metals, many more compounds are seen in the spectra of the red stars, mainly carbon compounds. Some compounds, observed usually only in red stars, have been detected in the sun during a total eclipse, their comparatively faint lines being no longer masked by the radiation from the whole surface, so we must not conclude that a star whose spectrum shows only the lines of helium, hydrogen, and oxygen, contains no other substances, but merely that these elements very largely predominate in its outer layers. There is every reason to believe that our yellow sun was formerly a much brighter white star, such as Sirius or Vega, and that in the course of ages it will cool down to a red star, such as Betelgeuse, with its heat emission diminished to about a seventh of its present amount. The earth would become a frozen void long before this happened.

When the sun's surface is examined through a powerful telescope, the features which first arrest the attention are the spots which were first observed by Galileo in the year 1610. It is very rarely that the visible surface of the sun is entirely free from these spots, though they vary greatly in number and magnitude, attaining a maximum value at intervals averaging about 11·1 years. They are not scattered all over the surface, but are confined to two belts in the neighbourhood of the sun's equator, and on either side of it. As Norman Lockyer has expressed it, the spots appear as though floating on the bright surface of the sun, which is known as the *photosphere*. They generally exhibit three shades of darkness, the darkness increasing towards the centre. Sometimes, however, the darker portions are eccentric, and very irregular in outline. When the spots are kept under observation, they are found to be undergoing continual change; and this may sometimes be observed in the course of an hour. Some of them exhibit obvious indications of cyclonic action.

The whole of the brighter portions appear to be coarsely mottled, consisting of bright patches separated by comparatively small intervals of much lower luminosity. The largest and brightest of the bright patches are observed in the immediate neighbourhood of the spots, and are known as *faculæ*. Some of these have been observed of such intense brightness as to indicate temperatures of from 10,000° to 12,000° C. The spectroscopic study of the sun shows that the spots are due to downrushes of gas, which, passing as they do from the cooler exterior of the sun to the hotter interior, do not undergo any condensation, which, if it occurred, would result in the production of clouds. The faculæ, on the other hand, are due to ascending masses of gas, which condense into clouds as they rise and spread over large

areas. These clouds consist, in all probability, mainly of liquid or solid particles of carbon and various metals.

Above the luminous clouds of the photosphere is the atmosphere of the sun, extending to a height estimated at between 5000 and 6000 miles. This atmosphere contains, in the form of gases, most of the elements found on the earth. Hydrogen and helium form a large proportion of the atmosphere, especially, owing to their relative lightness, of the upper layers. In the course of an eclipse this atmosphere becomes visible as soon as the luminous clouds of the photosphere have been hidden by the moon, and as it usually, under such circumstances, exhibits the purple coloration characteristic of hydrogen, it is known as the *chromosphere*. Flame-like masses, whose appearance has been compared to that of blades of grass, are seen to rise above the chromosphere all over its surface, being larger and more numerous when the sun-spots are at their maximum. They are described as solar prominences when they attain a height approaching 10,000 miles and upwards. Prominences have been observed some 370,000 miles in height. Norman Lockyer and Janssen showed in 1868 that their investigation could be continued, by the aid of the spectroscope, quite irrespective of eclipses. Those in which the most violent action is observed, indicated by the rapidity with which they change their size and shape, are found to consist largely of metals in the gaseous state. Movement at a speed approaching 300 miles a second has been observed in some of these metallic prominences. They appear only within the sun-spot area. The greater prominences are found to consist almost exclusively of hydrogen and helium, and they appear at all parts of the sun's surface. Their appearance is very like that of clouds, or masses of smoke, and prominences occurring near the poles have sometimes been kept under observation during a complete rotation of the sun, occupying about forty days, and even for longer periods. Arrhenius points out that the higher portions of some of these prominences must be almost in a vacuum, so that they cannot be supported by surrounding gases like the drops of water in terrestrial clouds. He explains their continued maintenance at such heights above the sun's surface by the repulsive effect of the radiation pressure which is considered in Chapter XII.

The question we have now to consider is the nature of the source of the energy required to project masses of metallic and other vapours from the sun's surface with velocities of some 300 miles a second, or about 600 times the maximum initial velocity of a rifle bullet.

The only conceivable source of this energy is the disintegration with explosive violence of some substance within the sun's interior.

If the energy so liberated could be applied to setting the gas in motion under as favourable conditions as it is applied to the bullet in a rifle, a very unlikely supposition, the solar explosive would have to be more than 300,000 times as powerful as the powder used in the rifle: for the energy required is proportional to the square of the velocity. Moreover, only a portion of the solar explosive is utilised in accelerating the mass of gas, probably a small proportion of the whole, as the remainder must be absorbed in the passage from the sun's interior to the surface. It is probable, therefore, that the energy developed in these solar explosions is at least a million times as great as could be produced by the strongest terrestrial explosives.

No such substance is known to us, so we must try by indirect means to ascertain its probable nature. In this we are greatly assisted by the fact that the materials of the sun are practically identical with those of which the earth is composed. The differences are therefore those of the same substances under different conditions of temperature. We may therefore assume that these high explosives in the sun are built up of materials with which we are familiar, and therefore the question we shall naturally put to ourselves in the first place is—Are these substances more likely to be what are known as chemical compounds, or are they radioactive bodies? Some of the former liberate energy when they disintegrate, others absorb it. The latter are, of course, out of the question. In favour of compounds of the former class is the fact that we are acquainted with numerous substances, known as explosives, which disintegrate with great suddenness and with a considerable liberation of energy. I am not aware of the existence of any explosive which disintegrates on lowering the temperature, as would be required in the case of the solar explosive. It is on raising the temperature that terrestrial explosives disintegrate. This, however, is a point of minor importance compared with the difficulty, amounting, I think, to impossibility, of imagining any combination of known elements which could liberate anything approaching the quantity of energy required by being merely resolved into its constituent molecules, unaccompanied by the atomic disintegration which is the distinguishing characteristic of radioactivity.

On the other hand, a substance such as radium, or one from which radium has been formed, would be perfectly capable of supplying the energy required, if the disintegration which, under present terrestrial conditions, takes place gradually, could be effected with sufficient suddenness.[1] Indirect evidence in favour

[1] Professor Arrhenius definitely suggests a similar conclusion in his recently published work (1909), *The Life of the Universe*, vol. ii. p. 239.

of the radioactive hypothesis is to be found in the presence of such large quantities of helium in the sun, and especially in the fact that when a sun-spot is spectroscopically examined, helium is invariably found to be present in quantity.

Our next step must be to endeavour to ascertain the conditions under which the formation and the disintegration, respectively, of radioactive substances is to be expected.

The radioactive elements known to us are all bodies of relatively high atomic weight. In other words, they are exceptionally closely packed bundles of matter. As far as it has been possible to obtain the various products of disintegration in sufficient quantity to permit of the atomic weight being determined, it is found that the atomic weights of successively formed products compose a steadily diminishing series, each consisting of a less closely packed parcel of matter than its immediate predecessor. The mechanical theory of heat teaches us that the building up of such a substance will be facilitated by increasing external pressure, and that diminution of pressure will, on the contrary, be favourable to further disintegration. Further, each step in the disintegration is accompanied by the liberation of energy in large quantities. The building up must, therefore, be accompanied by a similar large absorption of energy at every stage. The mechanical theory of heat shows that the formation of such substances is to be looked for under conditions of rising temperature. The applicability of these considerations to the problem of the origin of the radioactive substances has been pointed out by Arrhenius, who infers that they are most probably formed during the earlier stages of the growth of a star when the pressure and the temperature are both increasing.[1] Long ages ago both the sun and the earth must have passed through this stage. They are now gradually falling in temperature, but the radioactive substances once formed within them retard the process by the gradual liberation of the enormous stores of energy which were absorbed in their formation. The continual liberation, in the form of heat, of these immense stores of energy will in all probability render the continuance of life possible on the earth for many millions, possibly thousands of millions, of years to come. Had it not been for them, there is every reason to suppose that the earth would, many ages ago, have fallen to a temperature too low for life to exist on its surface, and it is quite possible that the sun itself might already have ceased to shine, and been transformed into a "dark star."

It was formerly supposed that all chemical compounds would be decomposed into their elements at the high temperatures existing

[1] See Chapter XX.

in the interior of the sun, but the spectroscope indicates the presence of such compounds in the lowest depths of the cyclonic vortices which we call sun-spots, while, on the other hand, comparatively few compounds appear to be capable of existing on, and immediately above, the surface of the photosphere. It may also be pointed out that many compounds are now known to chemists which can only be artificially produced in our laboratories at the temperature of the electric arc, which is considerably lower than that of the surface of the photosphere, and almost infinitesimally small in comparison to the temperatures which are probably to be found in the central portions of the sun. This formation of chemical compounds at high temperatures, though pointing in the same direction as the conclusions otherwise arrived at with regard to the origin of radioactive substances, cannot be regarded as affording very definite additional evidence in their favour. Chemical compounds are formed by the building up of the atoms of various elements into molecules without any destruction and rearrangement of the components of the atoms themselves, or the sub-atoms, as we have called them. Radioactive substances, on the other hand, are certainly built up by means of rearrangements of the sub-atoms into new elements, and it might be supposed that so fundamental a distinction would render valueless any analogies based on assumed similarities in their methods of formation. There are, however, very strong grounds for believing that the rearrangement of the sub-atoms of a radioactive substance, during transmutation into a higher or lower member of its series, is in general a very partial one, in which many of the groups of sub-atoms remain unchanged. This question is considered in Chapter XXI., and it will there be seen that confirmatory evidence obtained from the formation of chemical compounds is by no means valueless.

There is another direction from which further evidence might possibly be forthcoming. It has been pointed out that terrestrial observations have so far afforded no indications of the explosive disintegration of radioactive substances. If radioactive substances are built up at very high temperatures and pressures, as is indicated by thermodynamic considerations, explosive disintegration, if it takes place at all, is most likely to occur in the case of a very unstable member of the radioactive series experiencing a great and sudden diminution in pressure and temperature. If these conditions concur anywhere on the earth's surface, it will be during a volcanic eruption. I believe that geologists are agreed that the explosive violence with which gases and liquid matter are ejected during these disturbances is fully explained by the enormous pressures arising from the absorption, by molten siliceous

rock, of water percolating to great depths, owing to its displacing silicic acid at temperatures above 300° C., and the large resulting expansion. Should it ever be found that radioactive products are liberated in any considerable quantity during such disturbances, it would afford valuable confirmation of the hypothesis which I have suggested as a possible method of accounting for the origin of the solar prominences, and it might in that case be inferred that some part of the explosive phenomena observed in connection with terrestrial volcanic action was probably also due to the disintegration of radioactive material.

CHAPTER XIX.

THE SOLAR CORONA, THE AURORA, AND COMETS' TAILS.

SHORTLY before Lebedeff's experimental verification (referred to in Chapter VIII.) of the existence of a radiation pressure, Arrhenius called attention to its importance in providing possible explanations of various celestial phenomena. Since then he has devoted much attention to this matter, and a very good summary of the results arrived at is given in popular form in a little book which was published at the end of 1907 and has since appeared in an English translation under the name of *Worlds in the Making*. The radiation pressure at the sun's surface is very much greater than its value at the surface of the earth. In the former case it is estimated that the pressure, arising from this source, on the surface of a black body on which the rays fall vertically, is equivalent to a weight of about two and three quarter milligrammes on every square centimetre of surface exposed. If we imagine a small black sphere reflecting the whole of the light falling upon it, then, if this light is all of the same wave-length, and consequently of the same colour, it is easy to calculate the total pressure on the sphere due to such "homogeneous" radiation. The direction of the pressure is the same as that of the rays impinging on the sphere. The radiation from the sun is not homogeneous, but if we take the wave-length as that of the maximum radiation, the result so obtained will be only very slightly in excess of the true value. Arrhenius has shown in this way that a little sphere of the kind supposed, and of the same density as water, would just be maintained in equilibrium near the surface of the sun under the attractive force of the sun's gravity and the repulsive force due to its radiation, if the diameter of the sphere were about one and a half thousandths of a millimetre. The repulsive force would attain a value about ten times as great as the attraction to the sun in the case of spheres with a diameter about a tenth of this. The total radiation pressure on a body is, other things being equal, proportional

to its effective area.[1] Now this area is proportional to the square of the linear dimensions, while the volume, and therefore the mass, is proportional to the cube. The effect of the radiation pressure, relatively to the gravitational attraction, will therefore increase as the size of the body diminishes. Now we saw in Chapter XII. that theory indicates no limit to the smallness of particles capable of causing scattering of the radiation incident on them, and that there is experimental evidence of such scattering in the case of particles as small as the molecules of air. The radiation pressure may therefore be expected to give rise to very high velocities when it acts on the alpha particles, or the still smaller beta particles.

The Sun's Corona and the Polar Lights.—During total eclipses of the sun an outer luminous envelope is observed surrounding the chromosphere. It presents an appearance of great brilliancy when the moon has completely hidden the latter from view, but as long as any portion of the chromosphere remains visible, its superior brightness renders the adjacent portions of the *corona*, as it is called, completely invisible. The inner portion of the corona, extending to a distance of only a few thousand miles beyond the chromosphere, may be described as a radiance or glare. Part of this glare has been shown to be due to the earth's atmosphere and other causes external to the sun, but spectroscopic examination shows that some of the light emanates from an outer solar envelope consisting largely of cool hydrogen, and containing, chiefly in its outer portions, an element unknown on the earth's surface, which has been called *coronarium*, and is apparently lighter than hydrogen. Far more interesting and puzzling are the streamers which form what is generally known as the outer corona, and which have been observed extending to distances of as much as 400,000 miles from the sun's surface. Spectroscopic observations, first made during the total eclipse of 1868, show that the streamers consist, not of gaseous matter, but of liquid or solid particles. During periods of minimum sun-spot disturbance the main streamers radiate outwards from the two equatorial points of the sun's disc, suggesting the existence of an equatorial belt of such particles ejected from the sun. The comparatively feeble streamers emanating from points nearer to the sun's poles appear to be bent round towards this equatorial belt, very much as though they were following the lines of force of a magnet with poles coinciding approximately with the sun's geographical poles. During periods of greater sun-spot disturbance, streamers appear to

[1] The effective area is here clearly the maximum sectional area perpendicular to the direction of incidence.

emanate largely from the two sun-spot belts. This sometimes results in the corona exhibiting a more or less quadrangular outline; or the streamers may even appear to radiate with approximate uniformity from all parts of the sun's disc.

Sir Oliver Lodge considers that these streamers may not improbably be due to the discharge of electrons from the sun into the approximate vacuum of its immediate neighbourhood, and also points out that this would account for the known fact that the earth is highly charged with negative electricity, as one of the consequences of this electronic bombardment, extending into regions of space where no residual matter exists, would be the gradual accumulation of negative electricity by the earth. He also points out that the current of electricity carried by these torrents of particles would afford an easy explanation of one class of magnetic storms which has long been known to be closely related to the variation of sun-spots.

"The electric nuclei, when they form ions, would also serve as centres for condensation of atmospheric water vapour at high altitudes, and so be liable to affect rainfall. Moreover, the fact that water vapour condenses more readily on negative than on positive ions seems to furnish us with one explanation of atmospheric electricity, for a fall of rain would bring down with it a negative charge, and would leave the upper regions positively electrified with respect to the earth's surface: and this agrees with the known sign of the normal field of electric force in the atmosphere."[1]

Arrhenius had already suggested that the streamers were composed of negatively electrified particles emitted from the sun, and made this hypothesis the basis for an explanation of the terrestrial phenomena commonly known as the aurora borealis, or the polar lights. The latter term is preferable, as the phenomena are equally marked in the vicinity of both the earth's poles, although they were first observed in the north polar regions owing to Arctic exploration having preceded expeditions to the Antarctic zone.

If the ejection of matter from the interior of the sun is, as we have seen reason to suppose, very largely due to the explosive disintegration of radioactive matter, the emission of both alpha and beta particles in considerable numbers is to be expected. The results of observations on the solar prominences, and more particularly the information which the spectroscope gives us regarding the materials of which they are composed, show, however, that the ejected material consists to a very great extent of ordinary matter projected by the force of the explosion. Indeed,

[1] *Electrons*, p. 168.

it is most probable that the more finely divided portions of expelled matter, generally known as "solar dust," bear a close resemblance to the clouds of "volcanic dust" ejected from terrestrial volcanoes. The scale of terrestrial eruptions is but a puny one in comparison with those observed in the sun, the greatest observed height of volcanic dust-clouds being only a trifle of some 18 miles, as compared with the heights of the solar prominences which have been observed to extend to as much as 370,000 miles. These comparatively insignificant terrestrial dust-clouds, however, frequently spread over wide areas through the action of atmospheric currents, and many readers will remember the remarkable red sunsets which were observed for weeks together after the eruption of Mont Pelée in the island of Martinique in the year 1902, and which were due to the dissemination of volcanic dust ejected during the eruption.

We may therefore consider the matter ejected from the sun as consisting mainly of particles of ordinary matter in a fine state of division, which we may call solar dust, accompanied, in all probability, by considerable numbers of the positively charged alpha particles and of the negative electrons or beta particles. An initial velocity somewhat less than 400 miles a second would carry any of these particles beyond the limits of the sun's gravitational attraction without the assistance of the radiation pressure. Now this radiation pressure alone, acting upon a sphere having a diameter of 15 microns, would carry it, if originally at rest, and of the same density as water, across the distance separating the sun from the earth, in about fifty-six hours at an average speed of about 500 miles a second. Similar particles of a less density than water would travel more rapidly, as also would smaller particles. A large proportion of the smaller particles of solar dust will therefore pass beyond the limits of the sun's influence and will ultimately fall upon the surface of other heavenly bodies. The velocities of emission of the alpha and beta particles respectively, as observed under terrestrial conditions, range up to values approaching 17,000 miles a second and 170,000 miles a second respectively. These values may possibly be somewhat different under solar conditions, but the probability of there being a difference sufficient to affect our present argument is so small as to be negligible. We shall therefore assume the values to be the same, and also that the penetrative power of the beta particles is about a hundred times that of the alpha particles, as is the case under terrestrial conditions. If they were emitted from the surface of the sun, they would have to traverse the whole thickness of its atmosphere, and in so doing, would lose the greater part of their initial velocity of projection, so that their speeds

would not greatly exceed those arising from the radiation pressure alone. It appears likely, however, that undisintegrated masses of the radio-active explosive would be thrown up to sufficient heights to allow the particles liberated by the further disintegration to pass beyond the influence of the sun's gravitational attraction at speeds comparable with those attained by alpha and beta particles under terrestrial conditions. When the ejected particles reach the outer and cooler portion of the sun's atmosphere, they will begin to increase in size by the accretion of moisture. Some of the more slowly moving particles may be expected, in this manner, to attain sufficient size for the effect of the radiation pressure to be more than counterbalanced by the sun's gravitational attraction, in which case they will remain suspended for a time in the form of clouds, and as they continue to increase in size by the further accretion of moisture, they will ultimately fall back upon the sun's surface. The ratio of the number of particles permanently ejected from the sun to the number falling back upon its surface will be higher in the case of the beta than in the case of the alpha particles, on account of the higher initial velocities of the former, and their smaller size; for, in consequence of their higher initial velocities, and their higher acceleration under the action of the radiation pressure, due to their smaller mass, the beta particles will be exposed for shorter periods than the alpha particles to the accretion of moisture. Moreover, owing to the smaller size of the former, the accretion of sufficient moisture to cause their return to the sun would require a far longer period of exposure than in the case of the latter, while the effect of the radiation pressure would be greater relatively to the gravitational effect. The proportion of uncharged solar dust would, for similar reasons, be still smaller than in the case of the alpha particles. The larger proportion of beta particles ejected, as compared with alpha particles, will give rise to a resultant flow of negative electricity from the sun, as is required for the explanation of the observed phenomena of terrestrial atmospheric electricity. Of the charged particles, it will follow, from experimental results considered in Chapter VII., that many will coalesce in the journey from the sun, and so increase the proportion of neutral, or uncharged, particles.

The observed deflection of the coronal streamers towards the sun's equator certainly suggests the presence of charged particles, and also the existence of a permanent solar magnetic field of a somewhat similar character to that of the earth. This would, as we know, deflect the beta particles much more than the alpha ones, but if this is the true explanation of the deflection observed,

the smallness of its amount shows that the solar magnetic field cannot be strong enough to sensibly diminish the proportion of particles escaping, but will merely result in somewhat increasing the concentration of the streamers towards the equatorial disc. We should expect this disc to consist of two comparatively dense portions corresponding to the sun-spot belts extending from latitude 30° to 35° on either side of the sun's equator.

It is suggested by Arrhenius that the polar lights are due to the impact of this solar dust upon the upper layers of the earth's atmosphere, and to the reproduction of vacuum tube phenomena in the highly rarefied air by the charged particles following the lines of force of the earth's magnetic field which converge towards two points a little distance below the earth's surface in the neighbourhood of the north and south poles respectively. If the polar lights are produced, as suggested, by the streamers from the sun's corona, we should expect to find a distinct minimum in the displays of polar lights when the earth is in the sun's equatorial plane, and a distinct maximum when it approaches most nearly to the latitudes of the sun-spot belts. Now the plane of the earth's orbit, or plane of the ecliptic, as it is also called, is inclined at an angle of about 7° to the sun's equatorial plane. The earth would therefore never appear in the zenith, or vertically overhead, to an observer on the sun's surface at a greater latitude than 7°. The earth crosses the plane of the sun's equator on the 4th of June and the 6th of December, and attains its greatest distance from it three months later, viz. in September and March respectively. The largest number of particles should therefore reach the earth during the two latter months, and the smallest number during the two former ones. These conclusions appear to be confirmed by observations of the relative frequency of polar light displays, considering the difficulty of obtaining continuous records. Within the polar circles, where the most consistent results are to be expected, the continuance of daylight makes observations practically impossible, and even in somewhat lower latitudes the twilight interferes greatly with the observations. Arrhenius and Ekholm have tabulated a large number of records in Sweden, Norway, and the United States, and a smaller number in the southern hemisphere. In the four sets there are three well-marked maxima in March, one in April, one in September, and three in October.

A. Paulsen, who has added considerably to our knowledge of the polar lights, finds that they fall naturally into two distinct classes. Those of the first class consist of a soft glow, usually appearing in the form of an arch, or series of arches, one above the

other, and having their apices in the magnetic meridian of the place of observation. This luminosity has been observed at heights of as much as 250 miles. They are almost exclusively confined to the neighbourhood of the polar regions, and their maximum frequency is attained, not at the poles, but in narrow circular belts which include the magnetic and geographical poles. The northern one passes to the north of Nova Zembla, along the north-western coast of Norway, a little to the south of Iceland and Greenland, across Hudson's Bay and the north-western portion of Alaska. The light from this form of the aurora is remarkably constant, and the arches do not undergo any sudden changes in form, but usually appear to drift slowly towards the zenith. It may be owing to this constancy that no direct connection has been observed between auroral displays of this character and terrestrial magnetic disturbances.

The second class of polar lights is formed of those which have the appearance of rays or streamers, which are sometimes quite separate, but more frequently melt into one another, especially in the lower portions, giving rise to the appearance of thin drapery hangings. Their direction is always approximately that of the earth's magnetic force at the place of observation. They undergo great and rapid variations which are observed to influence the magnetic needle. It was noted by Paulsen that as one of these ribbons of light passes through the zenith from north to south the magnetic needle is deflected towards the west. This, as he pointed out, indicates the downward passage of negative electricity such as would be due to cathode, or beta, particles. Auroras of this character are much more widely distributed than those of the first class, but they occur most frequently within the same belts as the latter. The periods of maximum and minimum frequency, respectively, of auroras of the second class are found to agree closely with those of maximum and minimum sun-spot frequency, and the corresponding periods of maximum and minimum variation in the direction and intensity of terrestrial magnetism, but no such relation has been observed in the case of auroras of the first class.

It is evident that phenomena of the kind described above could not be due simply to uncharged particles of solar dust, which would be unaffected by the earth's magnetic force. We may therefore leave the uncharged particles out of account, and concentrate our attention entirely on the charged ones. As these approach the earth's surface they will enter the earth's magnetic field. Now, most of the earth's lines of force leave and enter the earth's surface within the northern and southern belts, respectively, of maximum auroral display. The earth's

magnetic field might, indeed, be represented, to a fairly close approximation, by that of a bar magnet slightly shorter than the earth's polar diameter, placed with its centre at the centre of the earth, and its ends vertically beneath the earth's magnetic poles. It will be seen, then, that the earth's lines of magnetic force will intersect the plane of the equator very nearly at right angles. The electrically charged particles which meet the magnetic lines at right angles will, therefore, merely be deflected in a direction very nearly east or west, while those whose motions have a component either northwards or southwards will continue moving towards the north or south as before, but will also begin to move in a helical curve having its axis in the direction of the lines of force, the helix being right-handed or left-handed according to the sign of the charge and the direction of motion along the lines of force. The distance of the particle from the axis of the helix in which it is moving is given by the expression $mv \sin \theta / e\mathrm{H}$, as stated on page 95. For different beta particles the ratio m/e remains constant, and the value of H is fixed by the position relative to the earth, and increases rapidly towards the poles. The path of a beta particle entering the earth's field at a given point will therefore depend on the value of $v \sin \theta$. The smaller this product, the more closely will the particle circle round the lines of force until it reaches the earth's surface in the neighbourhood of one of the poles. That is to say, the smaller the velocity with which the particle enters a given portion of the earth's field, and the more acute the angle at which it meets the lines of force, the more closely will it follow these lines of force to the poles. If $v \sin \theta$ is made larger, either by increasing the velocity or by making the angle θ more nearly a right angle, the particle will move in a helix of larger diameter, and will therefore reach the earth at a greater distance from the poles. The same reasoning applies to the alpha particles when compared only with each other, but, as we saw in Chapter VII., m/e is very much larger for the alpha particles, so that they follow the lines of force in very much larger helices than the beta particles, and will therefore be less concentrated towards the poles.

It is clear that the number of particles from the sun which enter the earth's magnetic field along any meridian will be a maximum when it is noon on that meridian, as the sun will then be nearest the zenith. The auroral displays should therefore attain a maximum shortly afterwards. This cannot be fully verified owing to their being invisible except during the hours of darkness, but it is in accordance with the observed fact that auroral displays are more frequent before midnight than afterwards.

If we consider the effect of the inclination of the earth's axis to the plane of the ecliptic, we shall see that the effect of the recession of either pole from the sun will be to increase the number of particles meeting the earth's lines of force at such angles as to carry them towards that pole, and therefore to increase the auroral displays during the winter and diminish them during the summer. It would scarcely be possible, however, to observe the small secondary maxima and minima, as, in addition to being masked by those previously considered, polar auroras would be invisible during the light summer nights.

It is suggested by Arrhenius that auroras of the first class are due to comparatively slow-moving charged particles which are nearly all deflected to the polar regions when they enter the earth's magnetic field, and that those of the second class are due to more rapidly moving particles which are emitted from the sun in larger proportions during the periods of high sun-spot activity. He suggests further that they are carried away from the sun, not by their velocities of projection, but by the action of the radiation pressure; but this conclusion is based on the erroneous assumption that the radiation pressure is ineffective on particles of less than a certain minimum size.[1]

We must therefore modify the hypothesis by assuming that the auroras of the first class are due to beta particles moving slowly in comparison with their original velocities of projection, but very much more rapidly than those ejected by the radiation pressure alone. Such particles would hardly remain long enough in the neighbourhood of water vapour for any considerable accretion to take place. Their collision with molecules of the gases of the atmosphere would account for the luminous phenomena observed, and the small value of the ratio m/e, in combination with their comparatively low velocities, would account for their deflection towards the polar regions of the earth. The more violent solar eruptions, associated mainly with periods of high sun-spot activity, might be expected to eject larger quantities of radioactive substances, and to greater heights. More rapidly moving beta particles would therefore reach the earth, and these would be accompanied by a certain proportion of alpha particles. The

[1] Professor Arrhenius, in a letter of 20th December 1909, maintains the correctness of his conclusions, referring to a paper by K. Schwarzschild (*Münchener Berichte*, vol. xxxi., 1901, p. 293), cited by him in *Worlds in the Making*, p. 97. The calculations in this paper, however, are based upon Maxwell's equations of electric force referred to axes moving through the ether. Larmor has shown (*Æther and Matter*, chap. vii.) that these equations are incorrect, which accounts for the discrepancy between Schwarzschild's conclusions and those derived from the analysis and experimental results of Lord Rayleigh (see p. 270).

rapidly moving beta particles would undergo much less deflection in the earth's magnetic field than the comparatively slow-moving ones responsible for the auroras of the first class, and this would account satisfactorily for the wider distribution of auroras of the second class. Their higher velocities would increase their magnetic action. Moreover, their magnetic effects would be much more easily observed, in consequence of the variations in the direction of their movements, than those of particles closely following the lines of force. An explanation of the streamers might possibly be derived from these considerations, although I do not see how the mere increase of velocity, and of variations in direction, would account for the production of striæ apparently analogous to those seen in vacuum tubes, but following the lines of magnetic force instead of the axis of the tube. It appears to me, however, that phenomena of this kind are exactly what might be expected to arise in the presence of alpha, as well as beta, particles. The former would be even less concentrated at the poles than the rapidly moving beta particles. They would circulate round the lines of force in the opposite direction to that of the beta particles, in helices of larger radius, and the drift across the lines of force and the motion along them would be at a slower rate. The collisions which would, under these circumstances, take place between the two sets of particles would appear to be quite sufficient to account for the varied and rapidly changing phenomena exhibited by the auroral streamers.

The continuous emission of negatively electrified particles from the sun would, unless compensated in some manner, result in a steadily increasing positive electrification of the sun which would ultimately prevent any further emission of negatively electrified particles. This charge would, however, result in drawing in from the surrounding space negatively electrified particles emitted from other heavenly bodies. Any increase in the sun's electric charge would increase its power of drawing in negative particles, until a condition of equilibrium was attained.

Radiation Pressure and Comets' Tails.—The repulsion by the sun of the matter of which the tails of comets are formed is another phenomenon which is explained by the action of the radiation pressure. The existence of this repulsion was demonstrated by Kepler in 1618, who ascribed it to the impact of the luminous rays from the sun, but on the erroneous hypothesis, then generally accepted, that rays of light were due to the emission of small corpuscles by luminous bodies. Observations made during the latter part of the nineteenth century gave values for the repulsive force varying, for different comets, from amounts only slightly in excess of the sun's gravitational attraction to amounts about forty times as great. Spectroscopic observations show that the tails

of comets are composed of solid particles, and not of masses of gas, and indicate that they consist largely of hydrocarbons.

The luminosity of the masses of solid particles forming the tails of comets, considering the low temperatures which must prevail except when they are in the neighbourhood of the sun, has long been a puzzle to astronomers. It may be very simply accounted for by the impact of the electrons which, as we have seen, are in all probability continually leaving and entering the sun in large numbers.

CHAPTER XX.

RADIOACTIVITY IN STARS AND NEBULÆ.

REFERENCE was made in Chapter XVIII. to the suggestion of Arrhenius that the origin of the radioactive substances was to be looked for in the earlier stages of the growth of a star, when the pressure and the temperature are increasing together. I propose now to consider the nature of the evidence which leads to the conclusion that our solar system has passed through a stage in which these conditions were fulfilled. The first scientifically constructed theory of the development of the solar system was the nebular hypothesis worked out by Laplace in the year 1776. General suggestions of a somewhat similar character had been made by the religious mystic Swedenborg in 1734, and somewhat later by the eminent philosopher Kant, neither of whom, however, succeeded in developing his suggestion into a theory consistent with physical laws. Laplace started with a gaseous mass filling, and extending far beyond, the present limits of the planetary system, and having an initial rotation about a central axis. It was necessary to assume this initial rotation in order to account for the rotation of the system, as it could not arise from the central forces of gravitation. Such a mass would gradually contract under the influence of its own gravitation, and, under these conditions, it follows from the known laws of dynamics that the speed of rotation must increase. When the speed of rotation reached a certain critical value the centrifugal acceleration of the outermost particles would balance the acceleration due to gravitational attraction towards the centre, with the result that a ring of the outer particles would be separated and left behind by the contracting mass inside. A similar process would be repeated again and again until the increasing density of the central portion and the decreasing distance from the centre caused such an increase in the gravitative force as to prevent the separation of further rings. It was supposed that the rings so formed would in general be broken up by disturbing forces and form a single planet, which would in its turn throw off further rings which

would, for the most part, break up in the same manner to form satellites, but the existence of Saturn's rings appeared to indicate the possibility of the rings continuing in existence as such. This theory accounted for the fact that the orbits of the planets and their satellites all lay, as far as was then known, very nearly in the plane of the ecliptic, and all revolved round the sun in the same direction as that in which the sun itself rotates. Clerk-Maxwell's investigation of the conditions of equilibrium of Saturn's rings showed that liquid or gaseous rings would be unstable, and that this would also apply to continuous solid rings unless certain extremely improbable conditions were fulfilled. Any such rings must therefore break up into separate masses, and if these varied greatly in size the larger ones would tend to draw the smaller ones to themselves, until finally a single moon was formed. The only practicable solution therefore was that the rings consisted of an immense number of small bodies all of about the same size. All this agreed very well with Laplace's theory. It appeared probable that the conditions requisite for the persistence of such rings were not likely to be fulfilled very frequently, and, moreover, they would not in any case be as persistent as the ordinary single moons, for the collisions between them would result in many of them falling into the central body, while others might coalesce with each other, giving a ring of widely separated bodies such as we find in the case of the belt of small planets known as the asteroids. One considerable difficulty was to account for the separation of rings of sufficient depth to suffice for the formation of the larger planets, and there were other minor difficulties; but in spite of these, the theory accounted fairly well for the facts known at the time that it was propounded, and formed a most valuable working hypothesis. The discovery of the two outermost planets of our system introduced difficulties of an altogether higher order. The orbits of both Uranus and Neptune are inclined to the plane of the ecliptic at very considerable angles, far too large to be accounted for by the small disturbing forces assumed to be the cause of the comparatively small deviations of the other planetary orbits. In addition to this, Neptune's moon moves in the direction opposed to that required by Laplace's theory, and so also do Saturn's outer moon, which was first observed by Pickering in 1898, and Jupiter's eighth moon, which was discovered by Melotte at Greenwich Observatory in January 1908. The existence of these retrograde motions indicates a much greater complexity than was assumed by Laplace in the conditions under which our solar system came into being. Laplace's theory can, therefore, at most, only be considered as a first approximation, as far as our solar system is

concerned. If, however, the stellar universe has had a definite beginning, Laplace's theory may very well represent the method in which the earliest stellar systems were formed. Our present knowledge of these systems points to the conclusion that our solar system must be considered as of very recent origin compared with the countless ages during which the transformations from nebulæ into suns and their planetary systems, and from suns and their systems into nebulæ, have been taking place.

We must therefore approach the problem from a different point of view, and endeavour, from observation of what is now going on, to discover what are the conditions under which such systems as our own are formed. This has been done by Arrhenius, and set forth in much fulness of detail in his recent work, *Worlds in the Making*, to which I cannot do better than refer such readers as may desire fuller information on these deeply interesting problems. The rest of this chapter is indeed mainly a summary of his argument, with occasional modifications and extensions.

Perhaps the easiest way to approach this problem will be to consider the probable future history of our own solar system. The sun is now gradually falling in temperature, and if it were not for the retardation of this cooling process, due to the stored-up energy of radioactive bodies, it would cease to be luminous in the course of a hundred million years at the outside.[1] The radioactive retardation might extend this into thousands of millions, perhaps more, but a billion years would appear to be quite an outside estimate of its possible life as a luminous star. We know that the sun is moving at a speed of about thirteen miles a second in the direction of the constellation Hercules, and its present average distance from the nearest stars towards which it is moving is of the order of some sixty billions of miles. Taking account of the number of these stars and the sun's diameter of some 850,000 miles, Arrhenius estimates that the most probable interval before a collision with one of them would be of the order of a hundred thousand billion years, supposing that the number of visible stars represented the total existing number. We have, however, seen that our sun would have become an extinct star long before this. Assuming it to continue luminous for a billion years, as suggested above, its future existence as an extinct star would be at least a hundred thousand times that during which it remained luminous. We infer, therefore, that we may assume the existence of at least a thousand extinct suns to every visible one, and this would reduce

[1] Lord Kelvin's *Baltimore Lectures*, p. 535.

our estimate of the probable interval before a collision to something of the order of a hundred billion years. These figures must only be regarded as guided guesses, and while it is of interest to obtain some idea of the magnitudes involved, their nearness, or otherwise, to the truth does not in any way affect the conclusion that such collisions must take place in the stellar universe. Direct observations on large numbers of stars have shown that they are moving through space, some more rapidly, and some probably more slowly, than the sun. These measurements have been made possible by means of the spectroscope, which enables us to determine the speed with which any luminous star is approaching or receding from the solar system. Under these circumstances collisions must take place more or less frequently, but this conclusion does not rest upon such inference alone.

Astronomers have long been puzzled by the occasional sudden blazing forth of a star of extraordinary brilliancy in positions where previously no star at all, or only a comparatively faint one, had been visible. After a short time these stars were seen to diminish in brilliancy until finally they again became faint or entirely invisible. Before the invention of the spectroscope it was impossible to ascertain the meaning of these phenomena, but the results of spectroscopic observations upon the more recent "new stars" leave little room for doubting that what has been witnessed in such cases is nothing else than the result of such a collision between two stars. A noteworthy instance is afforded by the new star observed by Anderson in the constellation Perseus on the morning of the 22nd of February 1901. Before considering this instance in detail, I would remind the reader that stars are classified into "magnitudes," as they are called, according to their relative brightness, the brightest ones being known as stars of the first magnitude. Stars of about two-fifths the brightness of those of the first magnitude are said to be of the second magnitude, so that a first magnitude star is two and a half times as bright as a second magnitude star; a second magnitude star two and a half times as bright as one of the third magnitude, and so on. The star Nova Persei, when first observed on the 22nd of February, appeared to be of the third magnitude, although on a photograph taken only twenty-eight hours earlier, and showing stars down to the twelfth magnitude, it was quite invisible. The star must therefore have increased to at least five thousand times its original brightness within the short interval of twenty-eight hours. On the following day, the 23rd of February, the new star appeared brighter than any star in the heavens with the exception of Sirius. On the 25th it had begun to diminish in brightness, but was still of the

first magnitude. By the 27th of February it had fallen to the second magnitude, by the 6th of March to the third, and by the 18th of March to the fourth. After this the brightness, while still, on the average, continuing to decrease, began to exhibit fluctuations in intensity, the fluctuations having at first a period of about three days, which steadily increased. By the 23rd of June the star had fallen to the sixth magnitude, and the fluctuations gradually diminished until after the period had increased to about ten days, and then became imperceptible, so that the decrease in brightness continued with increasing uniformity. By October it had fallen to the seventh magnitude, by February 1902 to the eighth, by July to the ninth, and by December 1902 to the tenth. Since then it has gradually fallen to the twelfth magnitude. When the star was at its brighest the light emitted was bluish-white, and before the fluctuations began to be observed it had changed, through yellow, to a reddish colour. During the fluctuations in brightness the colour varied from whitish-yellow at the maximum intensity to reddish at the minimum. Finally, the colour gradually changed to pure white. Early spectroscopic observations showed that masses of comparatively cool hydrogen mingled with metallic vapours were being ejected from the star at a speed of some 450 miles a second; but at the beginning these masses of gas did not completely envelop the new star, as was shown by the continuous spectrum due to the liquid or solid hot central body, together with the dark lines due to the absorption of part of its light by the masses of gas. This continuous spectrum gradually disappeared, together with the lines due to the various metals at first observed. The hydrogen lines remained as distinct as ever, accompanied by lines due to helium, and the lines of the gas nebulium gradually became visible. This substance has not yet been discovered on the earth, but is very characteristic of the nebulæ, where it occurs in the gaseous state in conjunction with hydrogen and helium. It appears to be heavier than helium, which is itself heavier than hydrogen, and it seems to be capable of existing in the gaseous state at temperatures nearly as low as in the cases of hydrogen and helium.

The rate of motion of the ejected gases observed immediately after the appearance of the star indicates an explosive violence greater even than has been observed in connection with the solar prominences. A feature of extraordinary interest consisted in the ejection from the star of two distinct streams of particles, one set travelling outwards from the star with about twice the velocity of the other set. These streams were first observed telescopically about a month after the appearance of the star, and it was found possible to follow their outward progress for nearly a year, owing

to the formation of luminous clouds wherever they traversed nebular matter. Some extensive nebulæ whose existence had not hitherto been suspected were discovered in this manner. This appeared to me to suggest that the particles were electrically charged, like the alpha and beta particles, and that the luminosity was due to their ionising action on the gaseous masses of nebular matter. The rates of motion were such that the distances traversed in 300 days would subtend angles of seven minutes and fourteen minutes of arc, respectively, to a terrestrial observer. The distance of the star is estimated to be such that its light would occupy about 120 years in its journey to the earth, which is equivalent to a distance of about 720 billion miles. The speeds of the particles would therefore be about 11,600 miles, and over 23,000 miles, a second, respectively, or about one-sixteenth and one-eighth of the velocity of light. These, it will be noticed, are nearly twenty-five times and fifty times, respectively, the highest observed speed of the ejected gases. They are, in fact, of the order to be anticipated if the collision were accompanied by explosive disintegration of radioactive substances stored within the colliding stars, which appears to be confirmatory of the hypothesis which I suggested would make it possible for alpha and beta particles to escape from the sun at speeds comparable with those observed under terrestrial conditions

If the more slowly moving group consisted of alpha particles, the speed of the other group would be too low for us to identify them with beta particles, and it is difficult to imagine any cause operating to reduce the speed of the latter without at the same time reducing the speed of the former in a still greater ratio.

The circumstances differ so greatly from any which can be observed under terrestrial conditions that no suggested hypothesis can be much more than a guess, but it appears to me a not improbable guess that if the alpha particles consist, as Rutherford's recent results conclusively prove, of helium atoms carrying positive charges, then this second group may have been composed of hydrogen atoms carrying the corresponding negative charges. Since the mass of a hydrogen atom is almost exactly half that of a helium atom, this would give the observed speed ratio of two to one.

An alternative hypothesis, which would appear from chemical considerations to have a greater *a priori* probability, would be that the faster-moving group consisted of hydrogen atoms carrying, like the helium atoms, positive charges. The corresponding negative charges might then be supposed to have escaped from the star in a still more rapidly moving cloud of electrons, or ordinary beta particles, unassociated with matter.

Such a cloud would certainly give rise to much less ionisation in the nebulous matter in the neighbourhood of the star than would be the case with the more massive particles associated with ordinary matter. It is, however, difficult to imagine that it would cause so little ionisation as to have entirely escaped detection, and for this reason the former guess seems to me to be the one which has most probability in its favour.

In any such collision between two stars the chances are enormously against their relative motion at the moment of impact being in the direction of the straight line joining their centres, and the collision which gave rise to the star Nova Persei was certainly not a direct impact, but an oblique one. In any collision which did not approximate to a direct impact, a considerable portion of the energy would be employed in giving a high speed of rotation to the new star, and in general this would be so great that any rotational motions of the separate stars previous to impact would exert no appreciable influence on the final result. The heat produced by such an impact would not be sufficient to account for the emission of gases at the speed observed immediately afterwards in the case of Nova Persei, without the accompanying explosive liberation of energy which I assumed to take place, owing to the bringing to the surface, and resulting disintegration, of radioactive substances. It has been calculated by A. Ritter that when two stars of equal size collide, after approaching each other from an indefinitely great distance, the heat due to the energy of impact would only be sufficient to produce an expansion of the combined volume to a little more than four times its original amount. The calculation is given in one of a series of remarkable papers on "The Height of the Atmosphere and the Constitution of Gaseous Worlds," which were published in Wiedemann's *Annalen*, beginning in 1878, and extending over a number of years. I shall assume, then, that such an internal explosion does take place at the moment of impact.

Arrhenius suggests that the shearing effect of the impact will result in the egress of the products of explosion taking place mainly in two parallel, oppositely directed, streams at right angles to the relative direction of motion of the two stars at the moment of impact. These streams of ejected matter would partake of the rotation arising from the collision. They would expand and, therefore, cool rapidly as they receded from the star. The increasing quantity of fine dust arising from the explosively ejected matter, and from fine particles driven away from the star by the radiation pressure, would obscure the new star to a continually increasing extent, and so gradually transform its white brilliancy, first to yellow, and then to reddish. When these

clouds were close to the star they would appear to envelop it completely. As the outer particles receded to greater distances, their places would be filled by further ejected matter, so that the obscuration would continue to increase, and the outer portions of the streams, becoming more widely separated, would produce a visibly greater effect when their denser portions were interposed between the star and the earth, and so give rise to the fluctuations observed; and these fluctuations would be of increasing period, owing to the decrease in the angular velocity of the dust particles as their distance from the star increased. The increasing quantity of ejected matter, forming a dust-laden atmosphere round the star, would explain the continued decrease in brightness. Its return towards a whitish colour may very probably have been due to the finest particles of dust having been driven away by the radiation pressure to considerable distances, where they would have less obscuring effect, owing to their distribution over a wider area, and, no doubt, in part also to the gradual aggregation of the dust into coarser particles. Ritter's calculation, referred to above, points to the conclusion that a large proportion of the matter of such a star will remain concentrated near the centre, and this is evidently in agreement with the phenomena which I have described in connection with the birth of the star Nova Persei, the gradual disappearance of the central body being clearly due to increasing obscuration by the growing masses of light gases and other nebulous matter. The ultimate disappearance of the central body in the nebulous cloud will complete its resemblance to other stellar nebulæ, and it will then show only the comparatively faint luminosity common to these bodies and which is, in all probability, due to the ionisation of the light gases, nebulium, helium, and hydrogen composing their outer layers, by the incursion of charged electric particles expelled either from the central mass or from other heavenly bodies.

The gradual decrease in the rate of rotation of the streams of gases and small particles issuing from the central body will tend to make them curl round into spirals surrounding the latter, and so we arrive at a satisfactory explanation of the spiral nebulæ. The rapid rotation of the central mass will cause it to assume a disc-like form, thicker at the central portions, and becoming thinner towards the outer circumference. A considerable portion of the ejected matter will probably escape entirely from the system, while the remainder will continue to revolve about it, the spirals being gradually transformed into nebulous rings surrounding the central mass. The ring nebulæ are therefore to be regarded as later stages in the growth of nebulæ which once had the spiral form. As the rate of ejection of matter from the central body

diminishes, the central body will again become visible, except in cases where the surrounding rings intervene between the star and the earth.

As long as the outer gaseous masses continue their expansion they will become steadily cooler. The rate of cooling will be accelerated by the escape of the most rapidly moving particles from the outer layers of the nebula, for the temperature of a gas or other body is simply due to the rapid motion of its molecules, and falls as the average speed of this motion is decreased. At the same time the small dust particles will, by collision with each other, gradually aggregate into the masses, which, when they are drawn within the earth's influence and fall on to its surface, are known as meteorites. They are of an extremely spongy texture, and present all the appearance of having been built up by the aggregation of small particles of cosmic dust. When a meteorite enters the earth's atmosphere it becomes incandescent from the heat generated by friction with the air, so that only comparatively large masses can reach the earth's surface without disintegration. The far more numerous smaller particles are entirely disintegrated in passing through the atmosphere. These meteorites, when formed in the nebular envelope of the new star, will, in addition to drawing in the smaller dust particles in their neighbourhood, act as centres of condensation for the gases, consisting mainly of hydrogen. As these aggregates increase in size, the gravitation of the central mass will overcome the radiation pressure, which, moreover, has been steadily decreasing, owing to the cooling of the central body by radiation. Swarms of meteorites will therefore fall towards the central body and will tend to compensate the heat lost by radiation.

Ultimately, the gaseous outer envelope of the nebula will cease to expand and will begin to contract. J. H. Lane showed in 1870 that, as contraction goes on, accompanied by the cooling of the outer layers, the temperature will begin to rise.[1] Condensation will be further assisted by the immigration of external matter. Further swarms of meteorites will be drawn into the nebula from outside, and these may be reinforced by larger heavenly bodies passing into the nebula. These would act as still more powerful centres of condensation, and I may point out that many interesting photographs of nebulæ and star clusters show great "rifts," as they are called, from which the nebulous matter appears to have been cleared away by the passage of stars in this way. As the temperature and pressure increased, the lines of nebulium and

[1] The problem has recently been treated in great detail by Lord Kelvin in a paper entitled "The Problem of a Spherical Gaseous Nebula," *Phil. Mag.*, vols. xv. and xvi., 1908.

helium, and some of the hydrogen lines, would tend to disappear from the spectrum of the nebula, and to be replaced by an increasing development of metallic lines, indicating the approach of the solar stage, and it is at this stage that we should expect the formation of the radioactive substances to begin, the hydrogen and helium, drawn in to the central body by the meteoric swarms, entering into combination with other substances under conditions of enormous temperature and pressure, and accompanied by the absorption of an immense amount of energy, corresponding to the amount which would be liberated by their disintegration. As the contraction continued the pressure would never cease increasing, but the temperature would attain a maximum value, after passing through which it would begin to decrease again, and this decrease would continue until the cycle began again with another collision. The star would in the meantime pass first from the white stage to the yellow, when it would be in a condition similar to the present state of the sun. It would then gradually turn red, and finally lose its luminosity and become an extinct sun, unless of course a second collision occurred before it had reached this stage. Such a nebula we may suppose to have been the origin of our own solar system.

Arrhenius suggests that the throwing off of rings from the central body might very possibly account for the formation of some of the smaller planets such as the asteroids, but that the most probable source of the larger planets would be the larger immigrating masses, increased by drawing in the smaller ones near which they passed. This suggestion would certainly obviate the difficulty, referred to earlier in the chapter, of accounting for the separation of rings of sufficient depth to suffice for the formation of the larger planets.

A far more probable suggestion, however, in my opinion, is that of Sir George Darwin, that the separation from a planet of material to form a satellite may have taken place in some cases by the breaking away of a single mass. A series of extremely powerful and elaborate mathematical investigations by Sir George Darwin and by H. Poincaré suggest that the first instability produced by cooling and shrinkage, with constant moment of momentum, in a rotating liquid planet might possibly give rise to a stable figure with a protrusion on one side of the centre of what had been the equatorial circle, and a flattening of the surface on the other side of the centre of inertia. This protrusion would increase with further cooling and shrinkage until, ultimately, the equilibrium would become unstable, and a comparatively small portion of the liquid would break away from the main mass.

If the whole planet were liquid at the time of separation, no

permanent scars would remain on either body, but if, as Lord Kelvin[1] suggests may have occurred in the case of the separation of the moon from the earth, it had attained the stage of a conglomeration of crystals with liquid lava filling the interstices between them, the plasticity would be sufficient to permit of the necessary changes of shape anterior to the separation, while allowing permanent scars to remain.

The scar left on the semi-plastic earth by the tearing away of the moon, and the subsequent surgings, might account for the persisting deviation of the figure of the earth from that of an exact figure of revolution, and also for the deviation from equilibrium between gravity and centrifugal force presented by the elevations forming the continents, and the depressions forming the oceans.

Sir George Darwin has worked out the subsequent stages passed through by the earth and moon, and accounted, in accordance with this theory, for the existing eccentricity of the moon's orbit. He has also shown that the earth's axis could not remain perpendicular to the orbital plane of the earth and moon, as such a system would be unstable, any deviation from exact perpendicularity being augmented by the earth's viscous resistance to change of shape.

Very large masses, or smaller masses moving with sufficiently high velocities, might pass through the nebula, and, after merely undergoing deflection by the central body, escape entirely from its influence. Smaller, or more slowly moving, masses would move round it in elliptic orbits. The friction between such a mass and the rotating nebulous matter would reduce the effect of the initial velocity of the mass compared with that due to the attraction of the central body. This would cause the elliptic orbit to approximate towards the circular shape, and the large space occupied by the central attracting body would aid in accelerating this change. The plane of the orbit would at the same time be deflected towards the equatorial plane of the central body. Whatever axial rotation the planetary body might initially possess, its axis would tend to become perpendicular to the plane of its orbit, and therefore perpendicular to the equatorial plane of the central body, with its axial rotation in the same direction as that of its revolution. These effects would be more powerful, the smaller the distance of the planetary mass from the central body, and large deviations might be expected in the case of the more distant masses. These planetary masses would in the same way

[1] "On the Formation of Concrete Matter from Atomic Origins," *Phil. Mag.*, vol. xv., 1908, p. 397. This paper contains a full list of references to the investigations of Darwin and Poincaré.

become surrounded by satellites formed from smaller bodies near which they passed, but which were not small enough or moving sufficiently slowly to be drawn into them. Similar considerations would apply to the orbits of these satellites. The planetary masses and their satellites would at first be surrounded by great globes, more or less flattened by their rotation, of nebulous matter. The satellites, being the smallest, would be the first to cool down, and then the smaller planets; the central body, on account of its vastly greater size, continuing to exist as a luminous sun for long ages after its planets and satellites had cooled to comparatively low temperatures.

As the nebula contracted from its greatest extent to a state resembling the present conditions of the solar system, the immigration of external bodies would become less and less frequent, and the conditions would become continually less favourable to their increase in size. Such bodies as did find their way in might therefore be expected to be comparatively small, and to move round the central body in orbits of varying elongation in planes bearing no special relation to the equatorial plane of the central body, or plane of the ecliptic, as it is called, in the case of our solar system. For, after the concentration of the nebulous matter into the planets and satellites, there would be no sensible frictional action tending to transform the elongated ellipses into circular orbits, and to turn their planes into the plane of the ecliptic. It will be seen that the system at which we have arrived resembles our present solar system practically in every detail, the latest immigrants forming the comets and meteoric swarms.

CHAPTER XXI.

ARRANGEMENT AND NUMBER OF ELECTRONS IN AN ATOM.

THERE are various ways in which an atom may be supposed to be built up of electric doublets in orbital motion about one another which, as we have seen, is the nature, indicated by the electron theory, of the constitution of an atom. One of these hypotheses has been sufficiently elaborated for its consequences to be compared with experiment, and has been shown capable of accounting for many of the principal known properties of material molecules. It will therefore be of interest to consider it in some detail, although it must not be regarded as an established theory of molecular constitution, but only as an attempt towards the development of such a theory.

The type of atom in question was first imagined by Lord Kelvin, and discussed by him in 1901 in a paper entitled "Æpinus Atomised." Suppose, for example, that an atom is formed of three such doublets, and in such a manner that the positive electrification is distributed uniformly over a sphere, while the three complementary negative electrons are included within the sphere. It can be shown that such an atom would be stable, with the three internal electrons at rest and forming the apices of an equilateral triangle having its centre at the centre of the sphere. The distance of each electron from the centre would have to be a little more than half the radius (the exact distance being ·57 times the radius). If the electrons were describing circular orbits round the centre of the sphere, their distance from the centre would increase with the speed of rotation. When this speed of rotation became sufficiently great, the electrons would at first begin to rotate outside the sphere, and would ultimately leave it altogether, when the atom would disintegrate.

Sir J. J. Thomson has investigated the conditions of stability of systems built up in this manner in a most able and interesting paper on the structure of the atom (*Phil. Mag.*, 1904). For the sake of simplicity in the investigation, the negative

electrons within the sphere were supposed to be confined to a single plane passing through the centre, and some very interesting results were arrived at. He found that two electrons would be arranged at equal distances on opposite sides of the centre, three would form an equilateral triangle as stated above, four would be placed at the corners of a square and five at the corners of a pentagon, or alternatively, four of them at the corners of a square and one of them at the centre. With six, seven, and eight electrons stability was also possible with one electron placed at the centre and the others arranged symmetrically about it. With nineteen, there would be an outer ring of twelve, an inner ring of six,

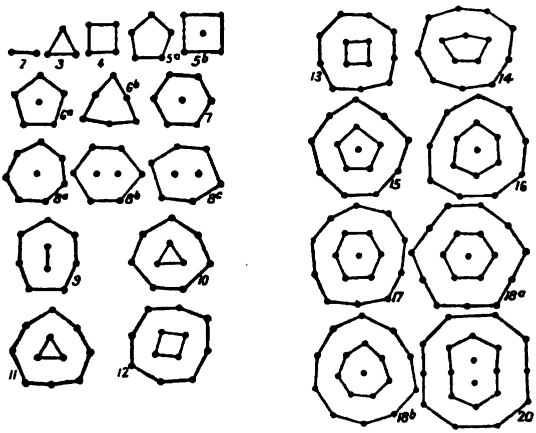


FIG. 31.

and one in the centre, and generally, as the number of electrons increased, the number of systems one within the other also increased. Some very beautiful experiments have been carried out by Professor Mayer which afford mechanical illustrations of model atoms of this kind with stationary electrons. The experiments were performed with a number of small magnets formed of equally magnetised steel needles made to float on water by being inserted into a small disc of cork, all the north poles, or all the south poles, being below the surface at the same time. The attractive force was provided by a stronger magnetic pole, opposite to that of the upward pointing poles of the small magnets, placed at some distance above the surface of the water. The configurations obtained with from two up to twenty magnets are shown in fig. 31.

If the electrons are not constrained to remain in one plane, many of these configurations will be unstable, unless the rings of electrons are in rotation in their own plane with a certain minimum angular velocity; for example, four electrons at rest will not be in stable equilibrium at the four corners of a square if there is any possibility of motion out of the plane, but will arrange themselves at the four angular points of a tetrahedron having its centre at the centre of the sphere. If, however, the ring of four electrons placed at the four corners of a square is made to rotate in its own plane with a certain minimum angular velocity, the configuration will become a stable one. The arrangement of five electrons at the corners of a pentagon will also be unstable when at rest, but stable when in rotation at a sufficient speed.

Thomson arrives at the conclusion that if the electrons are free to move in all directions they will arrange themselves in a series of concentric shells. The mathematical difficulties will then become very much greater, and he has not so far succeeded in obtaining a general solution for a distribution in shells, but it can be seen that the same kind of properties will be associated with the shells as with the rings, so he confines himself to the investigation of the simpler problem, and some of the results obtained are given in Table I., which is taken from Thomson's paper. Although this table only covers a very small range, it is sufficient to indicate the recurrence of certain groups at periodic intervals.

TABLE I.

Number of Electrons.	60.	55.	50.	45.	40.	35.	30.	25.	20.	15.	10.	5.
Number in successive rings	20	19	18	17	16	16	15	13	12	10	8	5
	16	16	15	14	13	12	10	9	7	5	2	
	13	12	11	10	8	6	5	3	1			
	8	7	5	4	3	1						
	3	1	1									

For example, we find that a single electron at the centre occurs in the case of a group of twenty electrons, again with thirty-five, fifty, and fifty-five, while a central ring of three electrons is found in an atom consisting of twenty-five, forty, or sixty electrons. Therefore, if one set of properties were connected with an atom having a single central electron, and another set with an atom having a central ring of three electrons, we should expect to find these two

kinds of properties periodically repeated with a series of atoms of different elements arranged in the order of the number of electrons contained in each. Now, the mass of an electron may be considered to be measured by the number of electrical doublets it contains, and therefore, since the electrons occur in pairs, by the number of electrons of one kind. A series of these hypothetical atoms, arranged in ascending order of the number of electrons contained in them, may therefore be considered as corresponding with a series of chemical elements arranged in ascending order of their atomic weights, or, more accurately speaking, their atomic masses.

Now, according to what is known as the *Periodic Law*, discovered by Mendeleeff, and independently by Newlands, if all the known chemical elements are arranged in the order of ascending atomic mass, they may be divided into a series of groups which are related among themselves in such a manner as would appear to indicate the periodic recurrence of certain similarities in structure of the atoms composing them. To illustrate this point further we will consider Table II., which is also taken from Thomson's paper,

TABLE II.

Number of Electrons.	59.	60.	61.	62.	63.	64.	65.	66.	67.
Number in successive rings .	20 16 13 8 2	20 16 13 8 3	20 16 13 9 3	20 17 13 9 3	20 17 13 10 3	20 17 13 10 4	20 17 14 10 4	20 17 14 10 5	20 17 15 10 5

referred to above, giving the entire series of atoms containing an outer ring of twenty electrons. Such an outer ring first occurs with an atom containing fifty-nine electrons, in which case the number inside is only just enough to make the outer ring stable. A small external disturbance will therefore be sufficient to detach a negative electron from such an atom, which will thus acquire a positive charge. It would not, however, be able to retain such a charge; for when it had lost an electron, the fifty-eight remaining electrons would arrange themselves with an outer ring of nineteen, which would be the last of the series containing nineteen electrons in the outer ring; that is to say, it would contain the maximum number of electrons inside, and would therefore be exceedingly stable, so that any further electrons would escape from it, while the positive charge on the atom due to the escape

of the fifty-ninth electron would immediately draw in a free electron from outside to replace the lost one. Such an atom would be neither electropositive nor electronegative, but would be incapable of permanently receiving any electric charge.

The atom containing sixty electrons could only lose one electron, and therefore acquire one electronic unit of positive electricity; for the loss of two electrons would reduce it to the system containing fifty-eight electrons, which would still more strongly attract external electrons than in the case last considered, owing to its acquiring a double charge. This system would therefore act like the atom of a monovalent electropositive element.

The atom containing sixty-one electrons, being more stable than the one with sixty, would part with its electrons less easily, but would be able to lose two before beginning to acquire fresh ones, and would therefore act like the atom of a divalent electropositive element. The atom with sixty-two electrons would part with its electrons still less readily, but would be capable of parting with three, and would therefore act like the atom of a trivalent electropositive element.

The atom containing sixty-seven electrons would have a very stable outer ring, but the addition of a single electron to this would lead to a rearrangement with an outer ring of twenty-one electrons, which has very little stability, and would consequently lose it again, so that in this case, as in the system of fifty-nine electrons, the atom would be incapable of receiving a permanent charge, and would act like the atom of an element having no valency.

The atom containing sixty-six electrons would be the most electronegative of the series, but would only be able to retain a charge of one electronic unit of negative electricity, for the acquisition of two units would transform it into a system of sixty-eight electrons. This atom would therefore act like that of a monovalent electronegative element.

Similarly, the atoms containing sixty-five and sixty-four electrons would behave like atoms of divalent and trivalent elements respectively. Thomson points out that such a sequence of properties is very like those of the elements given in the two vertical columns which follow, in which the elements are arranged in ascending order of atomic mass, the element of lowest atomic mass being at the head of the column, and that of highest atomic mass at the foot.

No valency	Helium	Neon.
Univalent electropositive	Lithium	Sodium.
Divalent electropositive	Glucinum	Magnesium.

Trivalent electropositive	Boron	Aluminium.
Tetravalent {Electropositive or Electronegative}	Carbon	Silicon.
Trivalent electronegative	Nitrogen	Phosphorus.
Divalent electronegative	Oxygen	Sulphur.
Univalent electronegative	Fluorine	Chlorine.
No valency	Neon	Argon.

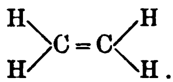
The central elements in these two series, viz. carbon and silicon, appear to be almost indifferently electropositive or electronegative, and this is just what might be expected according to the theory of structure which we have been considering.

If a number of electronegative atoms, in which the electrons are very firmly retained, are mixed with electropositive ones, in which the electrons are much less firmly held, it is to be anticipated that the mutual actions of the atoms may probably result in the transference of electrons from the electropositive atoms to the electronegative ones. The former would then become positively charged, and the latter negatively, when attraction would take place between the oppositely charged atoms, with the probable result of a compound being formed. Suppose, for example, that atoms of the divalent electronegative element oxygen are mixed with atoms of the monovalent electropositive element hydrogen, then, if each atom of hydrogen loses one electron and each atom of oxygen gains two electrons, the compound, water, H_2O, will be formed. This is the compound which is usually formed when these elements are mixed under favourable circumstances. Under special conditions another compound, HO (hydroxyl), is formed, and this can be simply accounted for by assuming that the circumstances are such that most of the oxygen atoms only receive one additional electron, so that the oxygen then behaves as a univalent element.

This variability in the valency of chemical elements under varying conditions may be further illustrated by the example of oxygen gas, which is ordinarily diatomic, that is, each of the molecules contains two atoms. One of these atoms must therefore be electropositive and the other electronegative, and it is easy, by comparing the position of oxygen in the series given above with its corresponding atom in Table II., to see that it might either lose one electron, and so become charged positively, or gain an electron, and thereby acquire a negative charge. In ozone, which consists of molecules of oxygen containing at least three atoms in each molecule, we have another example of such variability. From our theory we should expect that atoms of the higher

valencies would be capable of also behaving like atoms of lower valency, for if an atom is capable of losing or gaining two or more electrons, we may naturally expect it to be capable of losing or gaining only one.

From this point of view the *valency bonds* of graphic chemical formulæ represent Faraday tubes extending from the electro-positive to the electronegative molecules. Consider, for example, the hydrocarbon known as ethylene or olefiant gas, C_2H_4, which chemical considerations show to be capable of representation by the graphic formula



Since each bond, on this view, represents the transfer of an electron from one molecule to the other, this would give rise to two geometrically possible isomeric compounds; that is to say, compounds exactly similar as regards their component molecules, but differing in some of their physical properties. In one of these the two Faraday tubes connecting the carbon molecules would be similarly, and in the other oppositely, directed. Such geometrically possible isomers will be capable of physical existence only, provided both forms are dynamically stable.

Among the external conditions which may determine the passage of an electron from one atom to another, it is probable that the relative sizes of the atoms play an important part, as it may be shown that, if two atoms of different sizes are in contact, there will be a tendency for the smaller atoms to draw electrons out of the larger ones.

That electrical interactions may exist between atoms that have not become electrified by the gain or loss of one or more electrons, may be illustrated by considering the ideally simplified case of an atom assumed to consist of a single negative electron at the centre of a sphere of uniform positive electrification. There will be no electric forces between two such spheres as long as there is no interpenetration of one by the other, for, at any point outside either atom, the effect of the positive electrification uniformly distributed over its spherical surface will exactly counterbalance the effect of the equal negative charge at the centre. If, however, two atoms overlap each other, the attraction of the central negative electron in the first on the overlapping portion of the positive surface distribution of the second will be greater than the repulsion due to the positive distribution of the first. The electrons will remain at the centres of their respective spheres until the centre of each sphere comes within the other, when the

electrons will be displaced from their central positions, and pushed apart, so as to remain outside the line joining the centres. If the two spheres, instead of being of the same size, differ considerably, Lord Kelvin has shown that the central electrons from the larger one will remain within the smaller one when they are separated, so that one will be positively, and the other negatively, charged. If they were of the same size, however, they would remain as before after separation, each atom being neutral.

These considerations show the possibility of an electrical binding together of unelectrified atoms, and consequently also of unelectrified molecules, and if the cohesion of solid bodies were not already accounted for by the action of the gravitational forces known to exist, its explanation might be sought in electrical action of this nature.

If the atoms were surrounded by a good conducting medium, the electrons which had been expelled from an atom would more readily escape from its neighbourhood, without being drawn back again by the attraction of its positive electrification. If a number of atoms were moving about at high speeds they would not so readily enter into combination as if they were moving more slowly, or were at rest, for under the former circumstances an atom might frequently come within the sphere of attraction of another in a condition to enter into combination with it, and then be carried out again before combination had taken place, if the two were moving rapidly in different directions. This will explain why it is that raising the temperature of a solution or gas frequently causes chemical decomposition, while recombination frequently takes place on the gas or solution being cooled again. We should not, however, expect the atoms of two gases to necessarily form more complex combinations when cooled to a sufficiently low temperature, as we learn from the kinetic theory of gases that lowering the temperature of a gas does not affect the internal energy of the atoms to any large extent. Sir J. J. Thomson calls this internal energy the corpuscular temperature, as he applies the term corpuscle to the particles which are usually known as electrons. Following his suggestion, we might therefore say that the lowering of the molecular temperature of a gas does not necessarily affect the electronic temperature, and, unless this is sufficiently low, in relation to other circumstances, no combination can take place between two atoms, for such a coalescence necessarily involves transformation of potential into kinetic energy, that is to say, increase in the total energy of motion of the electrons, and if this energy were too great, a stable system would not be formed.

Since the molecular and electronic temperatures of gases are practically independent of each other, it follows that the energy lost when the motion of an electron is stopped by entering an atom is not lost directly by the collision, but by the resulting radiation.

Sir J. J. Thomson has calculated the loss of energy due to radiation for various systems of revolving electrons, and has shown that this loss diminishes very rapidly as the number of electrons in the system increases; for example, he finds that the radiation from each of a group of six electrons moving in a circle with one-tenth the velocity of light is less than one five-millionth part of the radiation from a single electron moving with the same speed in the same circle. At lower velocities the diminution is very much greater, *e.g.* the radiation from each of the six electrons would be only one fifty-thousand-billionth part of that from a single one, if the speed in each case were one-hundredth of that of light. The amount of energy involved in the formation of a complex atom, or in any rearrangement of the electrons within it, is extremely large; for example, the energy contained in a single gramme of hydrogen, on the assumption that each atom consists of 1000 electrons, would be sufficient to lift a weight of more than three hundred millions of tons to a height of a foot, or of generating 150 horse-power for more than an hour. This energy is proportional to the number of electrons in each atom, so that the amount stored up in a gramme of any other material will be greater than this in the same ratio in which its atomic mass is greater than that of hydrogen. When we consider that, while atoms are continually losing energy by radiation, they are at the same time receiving energy by radiation from other atoms, we can see how it is possible for such an atom to continue in existence, practically unchanged, for millions of years, as must have been the case.

The extent to which the foregoing representation of the internal structure of atoms accounts for the known physical and chemical properties of material molecules affords strong indications that it really forms a step, and an important step, in the direction of a consistent theory of atomic structure. The least satisfactory feature of the representation is the vagueness of the conception of the positive electrons forming a uniform electric distribution over a sphere enormously larger than the aggregate volume of the negative electrons, practically representing the boundary of the atom. It is not easy to see, indeed, how such a conception could be reconciled with the definite structure which Larmor assigns to an electric doublet, and which has proved sufficient over so wide a range of physical phenomena. It is, in

fact, not really a physical concept at all, but merely a provisional mathematical concept enabling the first steps to be taken gradually, by successive approximations, proceeding by the method of trial and error, and replacing the vague provisional concept by a definite physical representation. It affords, indeed, a very good illustration of the manner in which physical theories are usually built up. First a crude model is formed, perhaps, as in this case, with some of the parts only vaguely outlined. The theory of this crude model is then developed as far as the powers of mathematical analysis will allow, and the results compared with observation. In this way the partially fashioned portions can be gradually worked into suitable shape, while portions which cannot be modified so as to agree with the observed facts have to be replaced by others, and the process begun again.

The first assumption made as to the mass of, say, the hydrogen atom was that it was practically determined by the sum of the masses of the negative electrons, the mass contributions of the positive electrons being assumed to be negligible. This assumption was justified as follows by Sir J. J. Thomson in *Electricity and Matter*, p. 93:—

"Let us then take as our primordial system an electrical doublet, with a negative corpuscle at one end and an equal positive charge at the other, the two ends being connected by lines of electric force which we suppose to have a material existence. For reasons which will appear later on, we shall suppose that the volume over which the positive electricity is spread is very much larger than the volume of the corpuscle. The lines of force will therefore be very much more condensed near the corpuscle than at any other part of the system, and therefore the quantity of ether bound by the lines of force, the mass of which we regard as the mass of the system, will be very much greater near the corpuscle than elsewhere. If, as we have supposed, the size of the corpuscle is very small compared with the size of the volume occupied by the positive electrification, the mass of the system will practically arise from the mass of bound ether close to the corpuscle; thus the mass of the system will be practically independent of the position of its positive end, and will be very approximately the mass of the corpuscles if alone in the field."

Now, the mass of an atom of hydrogen is found to be about 1700 times as great as that of an electron, as represented by a cathode ray, or beta, particle, and therefore it would follow that a hydrogen atom must contain about 1700 electrons.

This conclusion appeared at the time to have nothing against it, and its truth was very generally assumed until Thomson's further investigations, which were published in the *Philosophical Magazine* for June 1906, showed very strong grounds for believing

it to be untenable. In this paper it is shown that three quite distinct lines of investigation lead to the extraordinary conclusion that the hydrogen atom contains only one active electron, and that the atoms of all other elements contain a number of active electrons comparable with their atomic masses reckoned on the basis that the atomic mass of hydrogen is unity. Commenting upon this in his book on *Electrons*,[1] Sir Oliver Lodge observes:—

> "This is an extraordinary and unexpected result, and at first sight appears very unlikely, since the ordinary chemical assumption of a unit atomic weight for hydrogen has always been known to be a pure convention, made for convenience alone, and not likely to correspond with anything in nature. I do not suppose anyone imagined that it would, even provisionally, be found to have a physical and rational basis."

The matter is of so much importance that I shall outline the nature of the arguments, referring for their fuller consideration to Chapter VII. of Sir J. J. Thomson's recent book on *The Corpuscular Theory of Matter*, where the question is considered in detail, or to the original paper.

The first argument is based on Barkla's observations on the Secondary Radiation produced by Röntgen rays. This arises from the sudden starting into motion of the electrons of molecules, those contained in the walls of the vacuum tube, for example, owing to the impact of the primary pulses. It can be shown that when the secondary radiation is of the same type as the primary, the pulses must be so thin as to sensibly act only upon one electron at a time, and the forces of restitution called into play by the displacement of the electron must be small in comparison with those exerted by the electric force in the primary impulse. Thomson has shown that when these conditions are fulfilled the energy of the secondary radiation emitted in unit time by unit volume of a space containing electrons is equal to

$$\frac{8\pi}{3}\frac{Ne^4}{m^2}$$

times the energy in the primary radiation passing through unit volume in unit time, where N is the number of electrons in the unit of volume, e the charge on an electron, and m its mass.

Barkla finds that for elements of small atomic weight the primary and secondary radiations are of the same type, and that this ratio of the energy of the secondary to that of the primary radiation is independent of the nature of the primary radiation

[1] *Electrons*, p. 151.

and directly proportional to the density, so that, where A is a constant,

$$\frac{8\pi}{3} \frac{Ne^4}{m^2} = A \times \text{density of substance}$$
$$= A \times \text{number of atoms in unit volume} \times \text{atomic mass.}$$

Therefore the number of electrons in an atom is proportional to the atomic mass of the substance.

In atmospheric air Barkla finds

$$\frac{8\pi}{3} \cdot \frac{Ne^4}{m^2} = \cdot 00025.$$

Now, in electromagnetic units,

$$e = 1 \cdot 2 \times 10^{-20}, \text{ and } \frac{e}{m} = 1 \cdot 7 \times 10^7;$$

so that

$$Ne = 10.$$

Now, if n is the number of molecules in a cubic centimetre of air at standard pressure and temperature 0° C, it is known that

$$ne = \cdot 4,$$

so that

$$N = 25n.$$

That is to say, the molecules of air contain on the average about twenty-five electrons, and, bearing in mind the proportions of oxygen and nitrogen in atmospheric air, and that their atomic masses are 16 and 28 respectively, the conclusion is arrived at that the number of electrons in the atoms of oxygen and nitrogen are 16 and 28 respectively, or are equal to their atomic masses.

Since the energy scattered by different gases traversed by Röntgen rays is proportional to the density, the number of electrons in unit volume must also be proportional to the density, so that if the above relation is true in one case it must hold generally. The atom of hydrogen, for example, must contain one *active* electron, for the above considerations would not apply to electrons strongly bound to the sphere of positive electrification, since the acceleration of such electrons would be insufficient to give rise to radiation of sensible amount in comparison with that arising from freely moving electrons.

The second argument is based on an elaborate calculation of the scattering effect of a material substance on cathode ray, or beta, particles traversing it. This argument leads to the conclusion that the number of electrons in an atom must be of the same order as the atomic mass, and would apply to bound as well as to free electrons. It is dependent, however, on the assumption that the repulsion between two electrons is always as the inverse square of the distance between them, however small that distance may be.

The third argument is based on the theory of dispersion of light in a monatomic gas, the atom being supposed to contain n electrons of mass m and charge e in a sphere of uniform positive electrification of mass M and charge E. Since the waves of light are very much longer than the diameter of an atom, the electric forces in the light wave may be regarded as sensibly constant throughout an atom. These forces will tend to increase the distances between the positive and negative charges, and, owing to the above-mentioned constancy, there will be time for a finite separation to be effected by each pulse. The result of this will be to reduce the velocity of the wave, giving rise to refraction. Now, the amount of separation will depend on the relative masses of the positive and negative charges, and if the mass of either were insensibly small compared to the other, there would be a complete shift at each alternation in place of a shift depending on the time. That is to say, the retardation would be independent of the wave-length, so that there would be refraction without dispersion. In the case of a monatomic gas the restoring force exerted by the positive charge on a displaced electron will be ν times the displacement, where ν is the index of refraction of the gas. This will not, however, be the case for a diatomic gas such as hydrogen, where the atoms are held together by the forces resulting from the displacement of the valency electrons. If the amount of dispersion were observed in the former case, an expression is obtained by Thomson in the paper referred to which would then give a relation between M and m.

It is there shown that if p is the frequency of the wave, ρ the density of the positive electrification, and N, as before, is the number of atoms in unit of volume,

$$\frac{\nu^2-1}{\nu^2+2}=\frac{\frac{4}{3}\pi\mathrm{N}(m\mathrm{E}^2+\mathrm{ME}e)}{\frac{4}{3}\pi\rho(\mathrm{M}e+m\mathrm{E})-m\mathrm{M}p^2}.$$

If λ is the wave-length, $p=2\pi c/\lambda$ where c is the constant of radiation, and p becomes null for infinitely long waves and, owing to the length of the waves in comparison with atomic dimensions,

the term mMp^2 will be a small quantity, so that, since $E = ne$, the equation may be written

$$\frac{\nu^2 - 1}{\nu^2 + 2} = \frac{NE}{\rho}\left\{1 + \frac{Mm}{Ee} \cdot \frac{3E}{4\pi\rho} \frac{1}{M + nm} p^2\right\}$$
$$= \frac{NE}{\rho}\left\{1 + \frac{m}{e^2} \cdot \frac{3E}{4\pi\rho} \cdot \frac{M}{n(M + nm)} p^2\right\}.$$

The only factor involving n being $\frac{M}{n(M + nm)}$, which is always less than $1/n$, the dispersion for an atom of given size will diminish rapidly as n increases, and measurements of the dispersion will enable the value of n to be estimated.

Now, experiments made by Lord Rayleigh appear to indicate that the dispersion in a monatomic gas is of the same order as in the case of diatomic gases. Results of the right order should therefore be obtained by using the value observed by Ketteler for the dispersion of the diatomic gas hydrogen, which he found to be expressed by the formula

$$\frac{\nu^2 - 1}{\nu^2 + 2} = \frac{1}{3}\left(2{\cdot}8014 \times 10^{-4} + 2 \times 10^{-14} \frac{1}{\lambda^2}\right).$$

Writing $p^2/4\pi^2 c^2$ for $1/\lambda^2$, and comparing the coefficients of p^2 in this expression and the preceding equation, we find

$$\frac{1}{n} \frac{M}{M + nm} = 1, \text{ approximately.}$$

This shows that n cannot differ much from unity, so that if a monatomic gas had the same density and optical properties as hydrogen it could not have many electrons in the atom.

The result obtained above also leads to the conclusion that M is very much larger than nm, so that instead of the mass of the positive electrons being insensible, in comparison with that of the negative ones, the positive electrons have much larger masses than the negative ones. We must therefore consider an electrical doublet as a very large sphere connected by a Faraday tube with a very small sphere, so that by far the greater part of the ether carried along by a tube is associated with the positive, and not with the negative, electron.

These results are extremely interesting, and at the same time extremely difficult to interpret, but they seem to show pretty definitely that the concept of a uniform distribution of the positive electrification over the surface of the atom is inadequate,

even as a mathematical concept, although it has been of value in making the first steps towards a theory of atomic structure.

Lodge (*Electrons*, p. 149) is of opinion that they tend to support the probability of the suggestion that the main bulk of the atom, in which electrons are embedded, may consist, not of positive electricity alone, but of a close admixture or combination of positive and negative electrons incapable of being appreciably separated by external forces,

"but are corporated together as a continuous mass, in the midst of which one or more isolated and individualised electrons may move about and carry on that display of external activity which confers upon the atom its observed properties."

He has pointed out that a structure of this kind would appear to afford a good basis from which to develop an electrical theory of cohesion.

"Consider," he observes (*Electrons*, p. 155), "the outer surface of an atom consisting of a regular group of interlaced electrons of alternately opposite sign. Its equipotential surfaces will be dimpled, or corrugated, or pimply sheets, which at a little distance away will be almost plane; but the dimples will increase rapidly in depth and become like the cover of a mattress, when something less than molecular distance—something approaching the internal electron distances apart—is reached. Two such atoms will therefore tend to settle down with their equipotential surfaces adjusted into uniformity, the pimples of the one fitting into the hollows of the other; and this is the state of things suggested by the facts of cohesion."

Sir J. J. Thomson's second argument would, it is true, conflict with this view, but this argument depends on the very doubtful assumption that the repulsion varies inversely as the square of the distance, however small the distance may be.

When $p=o$ in the formula on the preceding page, that is to say, for infinitely long waves,

$$\frac{\nu^2-1}{\nu^2+2}=\frac{\mathrm{NE}}{\rho}$$

$$= \mathrm{N} \times \text{volume of the sphere of positive electrification.}$$

Now, in a gas ν is so nearly equal to unity that

$$\frac{\nu^2-1}{\nu^2+2}=\frac{2}{3}(\nu-1) \text{ sensibly},$$

so that for gases $\nu-1$ will be proportional to the volume of the sphere of positive electrification. For the lighter elements the ratio of $\nu-1$ to the atomic weight is very irregular, as is shown

in the following table, due to Cuthbertson and Metcalfe (*Phil. Trans.*, A, 1907, p. 138):—

Gas.	$\nu - 1$.	Atomic mass.	$\frac{10^6(\nu-1)}{\text{Atomic mass}}$.
Helium	72×10^{-6}	4	18
Neon	137×10^{-6}	20	6·85
Argon	508×10^{-6}	40	12·7
Krypton	850×10^{-6}	80	10·6
Xenon	1378×10^{-6}	128	10·7
Mercury	1866×10^{-6}	200	9·3
Hydrogen	139×10^{-6}	1	139
Nitrogen	297×10^{-6}	14	21
Phosphorus	1197×10^{-6}	31	39
Arsenic	1550×10^{-6}	75	20
Oxygen	270×10^{-6}	16	17
Sulphur	1101×10^{-6}	32	34
Selenium	1565×10^{-6}	79	20
Tellurium	2495×10^{-6}	127	20
Zinc	2060×10^{-6}	65	30
Cadmium	2675×10^{-6}	112	24

but for those of large atomic mass in one group it is fairly constant, showing that the volume of the sphere of positive electrification is roughly proportional to the atomic mass when the atom contains many electrons.

I will conclude this chapter with the further consideration of the difficulty arising from the linear relations of the lines in the spectrum of a disturbed molecule, which was discussed at the end of the preceding chapter. We can now see that a collision between two atoms of the Thomson type may possibly give rise to effects of three distinct kinds, according to the intensity of the shock.

A moderate shock may be expected to give rise only to perturbed vibrations due to motion of the centre of gravity, as suggested by Lodge, and discussed in the preceding chapter.

A stronger shock will probably give rise to similar effects of a more intense character, and in addition, cause the expulsion of one or more of the mobile, or active, electrons contained in the atom, without otherwise altering its constitution. That is to say, it will, in addition to the effect in the first case, give rise to ionisation. A very violent shock might be imagined to be capable of causing sufficient disturbance to displace some of the permanently revolving electrons beyond the sphere of positive electrification, causing the disintegration of the atom.

The suggestion that ionisation may be produced in this way

is confirmed by observations by Stark, who has devoted a great deal of attention to the investigation of atomic constitution from the spectroscopic point of view. His own conclusion from his observations is that spectral lines are entirely due to ionisation. He observed in the first place that band spectra, that is to say, spectra in which the lines are widened out into comparatively broad bands, are produced by the neutral atoms, while line spectra derive their origin from atoms which have lost or gained one or more electrons. This would account for the increased complexity of spectra at high temperatures. He found that the periods of vibration, as shown by the lines of the spectrum of an electric discharge, are different according as they are looked at in the direction of the current, or at right angles to it; and the difference was exactly what it would be if due to the Doppler effect on the light emitted by bodies moving with the known velocities of the positive particles in the electric discharge. The conclusion that the production of spectra is *entirely* due to ionisation, even if interpreted in the sense of being invariably accompanied by ionisation, does not appear to be warranted by his experimental results. If the vibrations were directly due only to the disturbance accompanying ionisation, to the exclusion of the cause first considered, there would be relations between the squares of the frequencies instead of between the frequencies themselves, except in the case of the vibrations emitted by isolated electrons.

With regard to the third possible case, I am not aware of any evidence indicating the occurrence of such atomic disintegration arising from such shocks as are possible under present terrestrial conditions, but evidence of such occurrences under solar and stellar conditions has been given in previous chapters.

CHAPTER XXII.

CHANGES IN THE ASPECT OF FUNDAMENTAL MECHANICAL PRINCIPLES.

THE present chapter has been directly suggested by Poincaré's discussion, in *La Valeur de la Science*, of the past, present, and future of mathematical physics. His remarks on the past history of the subject are made from the same point of view as that developed in Chapter III., regarding it as originating in the theory of central forces as developed in Newton's *Principia* and further elaborated in Laplace's *Mécanique Céleste*. The motions of the heavenly bodies were found to be capable of very complete co-ordination on the basis of regarding them as bodies of such small size in comparison with their relative distances that the positions occupied by them might be treated as mere mathematical points, and their mutual actions expressed as attractions inversely proportional to the distances separating these points. When other physical phenomena were investigated, and were not found to be explicable on the basis of forces, attractive or repulsive, varying as the inverse square of the distance, the whole problem was assumed to consist in the determination of the exponent required to replace -2 in the law of the inverse square. The material universe was, in fact, regarded as made up of particles supposed always to be separated by distances which were large in comparison with the dimensions of the particles themselves, and always attracting or repelling one another according to some power of the distance, fractional powers being tried when integral powers could not be found which would lead to results in accordance with observations. This point of view was extremely fruitful, and led to the development of mathematical theories which not only made it possible to co-ordinate numerous phenomena but to predict many which were before unknown, and which were experimentally verified when searched for. It is therefore not to be wondered at that it should have come to be almost universally believed that this mechanical model, as we should describe it, really formed a complete representation of the existing

material universe, and that the search for deeper lying causes than forces varying according to powers of distances was a quest beyond the reach of the human intellect. There were a few, Newton himself being the most prominent, who were not satisfied with such a representation, and saw the necessity for an underlying ether as a basis for the mutual interactions of material bodies. To most physicists of those days the name of ether suggested only the pseudo-philosophical series of ethers and vapours which had been so plentifully employed as a shroud for ignorance and superstition. They looked upon it as a mere miasmatic fog which the light of scientific investigation was to dispel for ever, revealing the mechanism of the universe in a form as clearly defined, and almost as limited, as a delicate piece of machinery fresh from the workshop of a master craftsman.

We have now, as the result of long experience, learned the lesson that the most perfect representations which we can form of the complex of natural phenomena are really but the crudest of models, and that although there is no fixed and clearly defined barrier within which our investigations are confined, yet the barriers are no less real and effective for their lack of definiteness, and are to be found in the ever-increasing complexity of mutual relations and interactions revealed by each fresh advance.

The early limited conceptions were, however, as Poincaré points out, of inestimable service in teaching us the inner meaning of what we call physical laws, as nothing else than constant relations between phenomena, as true at one time as at another; in other words, differential equations. It soon became evident that the mechanism of nature was too complex to be fully investigated by the direct step-by-step process of the *Principia* and the *Mécanique Céleste*, and various mechanical principles were developed, enabling previously obtained results to be applied to new phenomena *en bloc* in place of attempting to follow them step by step in all their details, starting with Newton's law of action and reaction. Poincaré enumerates the most important of these as follows:—

(1) Newton's principle that action and reaction are equal and opposite.

(2) The law of the conservation of energy.

(3) The principle of relativity, according to which the laws of physical phenomena must be the same for an observer remaining in a fixed position as for one carried through space at a uniform speed without change of direction.

(4) The principle of the constancy, or conservation, of matter, which Poincaré quotes as the principle of Lavoisier.

(5) The principle of least action.

(6) Carnot's principle, or the principle of the dissipation of energy.

The principle of relativity was considered in detail in Chapter V., and Poincaré's views regarding it were discussed, so that it need not be referred to again. His views regarding the other principles enumerated are in one respect of a similar character. His attitude is one of apprehension that principles which have proved trustworthy guides up to a certain stage, appear to lose much of their definiteness, and therefore of their reliability as guides, as we penetrate more deeply into the mysteries of the universe of which we form a part. It appears to be inevitable that this should be so. All these principles are generalisations based ultimately upon observation, experimentally directed or otherwise. As the field of observation widens and increases in complexity our generalisations also extend their scope, and it becomes increasingly difficult to make sure that no essential circumstances have been omitted. If the central force model of the universe were a fully adequate representation, the principles of the conservation of energy and of least action might be regarded merely as useful theorems derived from the law of action and reaction. Even when this system held the field, however, it was necessary to recognise irreversible actions, such as frictional resistance, which could not exist in a universe regulated entirely by central forces.

This recognition was first definitely formulated in Carnot's principle, which in its original form, based on the assumption of the materiality of heat, simply expressed the fact that when heat was allowed to flow down a slope as it were, or gradient, from a hot to a cold body, it could be made to do mechanical work, but that it was theoretically impossible, by an equal expenditure of work, to raise the same quantity of heat up to the level, so to speak, of the hot body, so that the process was an irreversible one. When the true nature of heat was discovered this became the principle of the dissipation of energy, or the second law of thermodynamics, which has been stated by Lord Kelvin in the form :—.

"It is impossible,* by means of inanimate material agency, to obtain mechanical work by cooling a body below the temperature of surrounding bodies." When work is done by allowing heat, that is to say, unorganised energy, to flow from a hot body to a cold one, it is impossible to transform the whole of the heat into mechanical work, that is to say, into organised energy. The whole of a quantity of organised energy can, however, be transformed into the form of unorganised energy, and it is just the same in natural processes, so that there is a tendency for energy to pass from the organised to the unorganised form.

There is no more firmly grounded generalisation from experience than this, although Maxwell's illustration of the sorting demon shows that it could be overcome by means of intelligent agency without infraction of the principle of the conservation of energy, that is to say, without leading to perpetual motion. We are not, of course, entitled to conclude absolutely from our own limited experience that the law of the dissipation of energy necessarily extends to the whole material universe; but we are entitled to draw the conclusion, subject to the proviso that the whole universe is subject to the same laws as the portion which we have been able to explore, and on any other assumption we can draw no conclusions at all with regard to the whole universe. If it is subject to the law of dissipation, and there are no counteracting processes unknown to us, then the material universe cannot be a conservative system, but must have had a beginning and be approaching towards an ultimate end in which all the energy will be uniformly distributed in the form of unorganised molecular motions.

It is difficult to suppose that the stellar universe can be running down in this manner like a clock. The more extensive our means of exploring the stellar universe become, the more does it appear as though it were going through vast series of cyclical transformations, showing no direct evidence of growth or decay as a whole.

Arrhenius suggests that there may exist in the nebulæ a mechanism giving rise to actions of a similar character to those attributed to Maxwell's sorting demons, and so tending to counteract the dissipation of energy; that is to say, that while in some parts of the universe there is a tendency to a passage of energy from organised to unorganised forms, there may in other parts be a tendency acting in the reverse direction. He thinks it possible that the radiation pressure may have part in this action, his argument being, in effect:—that our solar system is at present in a certain stage of one of the great series of cyclical changes from nebulæ to star systems and from star systems to nebulæ, changes in which the radiation pressure plays, as we have seen, an important part, and that we have no right to argue, from the conditions existing in our solar system at the present time, as to what may be the conditions at far-distant periods of the cycle. A speculation of this kind might have been considered a legitimate one when the principle of the dissipation of energy was not known to have any other foundation than our direct experience within a limited portion of the universe. The demonstration in Appendix K that it is a consequence of the molecular constitution of matter, makes it inadmissible unless we extend

our consideration beyond the molecular scheme to the directive intelligence which rules the universe.[1]

Maxwell has indicated what appears to be the only way out of the difficulty. His illustration affords a perfectly definite and irrefutable proof that the dissipation is not, even according to our own limited experience, a necessary natural law in an intelligently directed universe, and the conception of a self-existent, undirected universe is one which becomes more and more untenable with every advance into the depths of Nature's mysteries. It is certainly absolutely irreconcilable with our present point of view, as is definitely set forth in Chapter XXIV.

Newton's law of action and reaction is absolutely unaffected, as regards its formal statement, by the passage to our present standpoint. Newton's original statement does not even formally restrict it to actions between material bodies, although its extension, now rendered necessary, to actions between ether and matter, were not, and could not have been, contemplated by him. The charge of vagueness and indefiniteness involved in extending this and other mechanical principles to the ether, which is strongly insisted on by Poincaré, is certainly not sustainable as regards Larmor's perfectly definite formulation of the electron; this criticism, indeed, refers definitely to Lorentz's electron theory, which was propounded in a much less definite and complete form, and assumed the absence of any mechanical interactions between the electrons and the ether, an assumption which is shown in Appendix H to be untenable.

The principle of conservation of matter is a purely experimental result, and it holds good to such very close approximation that for all practical purposes it is just as safe a guide as ever it was. No amount of experimental evidence can possibly afford grounds for postulating theoretical exactitude with regard to any statement respecting physical phenomena, for it cannot by any possibility tell us more than that the statement holds good within the limits of errors of observation. Here, in fact, the electron theory provides a new principle, the conservation of electrons, which is practically an exact statement, with a definite theoretical basis, the impossibility of creating or annihilating an electron

[1] The second law of thermodynamics is not shown, in Appendix K, to be quantitatively true, so that the illegitimacy of applying this law to systems of ether and matter, which was pointed out in Chapter XIII., still holds good. It is shown to be infinitely improbable, in the absence of intelligent purposive control, that the dissipation should vanish. It might, however, be either of greater or less amount, in the universe as a whole, than is indicated by the second law of thermodynamics, which is an empirical law known to hold, within the limits of experimental error, only in that portion of the universe which is accessible to our observation of its action.

by physical means. Yet even here the exactness is not absolute in the theoretical sense, for it is theoretically possible, although almost infinitely improbable, for two electrons to collide in such a manner as to annihilate each other.

The concept of mass is one which, from the theoretical point of view, is fundamentally changed, but from the ordinary practical point of view not at all, since changes in mass become sensible only when the body is in motion with a speed approaching that of light.

There are two properties associated with the mass of a body, viz. its weight under given circumstances, and its inertia. The weight is quite unaffected, as far as we know,[1] by our present point of view, but the inertia, which was formerly supposed to be invariable, is now shown to have this property only when the body is not in motion with a speed approaching that of light. When it is moving with such a velocity, the inertia will vary, not only with the speed, but with the direction of acceleration relative to the line of motion, as Abraham has shown. Moreover, as Searle pointed out in 1897, three methods of expressing the inertia, which are usually equivalent, viz. the ratio of force to acceleration, of momentum to velocity, and of the kinetic energy to half the square of the velocity, all lead to different results if the body is moving with a speed approaching that of light.

The principle of the conservation of energy certainly appears to assume a very much less fundamental aspect from the point of view of molecular physics, which I cannot illustrate better than by quoting Larmor's pronouncement (*Æther and Matter*, p. 286):—

"One effect of admitting a molecular synthesis of dynamical principles such as the one here described is to depose the conception of energy from the fundamental or absolute status that is sometimes assigned to it; if a molecular constitution is fundamental, energy cannot also be so. It has appeared that we can know nothing about the aggregrate or total energy of the molecules of a material system except that its numerical value is diminished in a definite manner when the system does mechanical work or loses heat. The definite amount of energy that plays so prominent a part in mechanical and physical theory is really the mechanically available energy, which is separated out from the aggregate energy by a mathematical process of averaging, in the course of the transition from the definite molecular system to the

[1] We shall see in the next chapter that certain outstanding anomalies in the motions of the heavenly bodies suggest the possibility of gravitational forces depending, to a small extent, on the motions of the attracting bodies, and that the most promising theories of gravitation appear to indicate the existence of such effects. They would, however, be far too minute to give rise to perceptible variations in the weights of terrestrial bodies.

material system considered as aggregated matter in bulk. This energy is definite, but is not, like matter itself, an entity that is concerned in unchanging amount; it merely possesses the statistical yet practically exact property, based on the partly uncoordinated character of molecular aggregation, that it cannot spontaneously increase, while it may and usually does diminish in the course of gradual physical changes."

If the dissipation of energy is compensated throughout the universe, then the principle of the conservation of energy will necessarily apply to the universe as a whole.

From the still more fundamental point of view set forth in Chapter XXIV. the principle of the conservation of energy will be seen to have a greater significance than it possesses even in the mechanical scheme. The energy of a dynamical system is there regarded as its sole ultimate physical content, so that the conservation of its energy becomes a necessary condition of its permanent existence.

The principle of least action has been fully considered in Chapter III., and Appendix B.

CHAPTER XXIII.

GRAVITATION AND COHESION.

THE law of gravitation formulated by Newton is an experimental law deduced from observations of the motions of the heavenly bodies. Newton's general statement of it, that every particle of matter in the universe attracts every other particle with a force proportional to the product of the masses of the two particles divided by the square of the distance between them, was expressly put forward as a hypothesis to be tested from the results involved in it. Its application to the motions of the heavenly bodies involved its correctness only for the attractions of bodies separated by very great distances in comparison with their sizes. In this respect there is certainly no physical law whose accuracy has been more firmly established. For a considerable period after it was stated by Newton it was regarded as open to doubt, as the advances in observational astronomy began to reveal irregularities in the planetary motions which at first appeared to be unaccounted for by it. Before long, however, the progress made by Laplace and others in the mathematical investigation of the motions of the heavenly bodies showed that these irregularities were due to the disturbing effects arising from the mutual actions between different planets and satellites, and every advance made in the mathematical investigation of the motions of the heavenly bodies afforded further confirmation of the exactness of Newton's law.

There are, however, some minor problems in the motions of the heavenly bodies which have not yet been completely solved, and which may possibly find their solution in Newton's law of gravitation not being an absolutely exact law, but only an exceedingly close approximation.

With the exception of gravitation and cohesion, the whole complex of material phenomena appears, so far, to be capable of representation in terms of actions transmitted through the ether. Cohesion, as we shall presently see, can be fully accounted for in terms of gravitational attraction acting at very small distances, so that the modern etheric scheme of the universe will embrace

the totality of material phenomena if only gravitation can be brought within its scope.

From this point of view we should expect Newton's law, like other physical laws, to prove to be ultimately an approximation founded upon the statistical uniformity, as regards our means of investigation, in the smallest portions which we can distinguish of the constituents of the material universe.

The most noteworthy discrepancies between the results of astronomical observation and those deduced from theory, on the assumption of the exactness of Newton's law, occur in:—the motion of the perihelion of Mercury, in which the discrepancy is about forty-one seconds of arc in a century; the motion of the perihelion of Mars; the motion of the nodes of the orbit of Venus. There also appear to be:—a small unexplained portion of the change in the eccentricity of the orbit of Mercury, some unexplained irregularities in the orbit of the moon, and some anomalies in the motion of Encke's comet.

The evidence of the exactness of Newton's law at moderate distances is all of an indirect character. The most convincing argument in its favour is the observed fact that the weight of a body does not sensibly vary with its direction, or with its position except as regards change of distance from other bodies. This shows that the transmission of the gravitational attraction between two bodies is not affected by intervening matter, and, on the assumption that the action is transmitted by the intervening medium, would lead to the conclusion that material molecules are widely separated from one another in comparison with their dimensions, a conclusion indicated, as we have seen, by other and independent considerations. Laplace has shown (*Mécanique Céleste*, livre xvi., § 6) that the attraction of a molecule at the centre of the earth upon a molecule at the surface could not, in accordance with the known agreement of the Lunar Theory with the results of observation, be diminished by as much as one part in a million owing to absorption in its passage through the earth. The only direct evidence of the law is the fair agreement with each other of measurements of the constant of gravitation under varying conditions, and they are not sufficient to establish the law of the inverse square with any considerable accuracy. The exactness with which the weight of a body is conserved throughout the most varied physical and chemical transformations shows that the mutual actions of neighbouring molecules cannot modify their structure in a way that affects their gravitational action, so that the separate particles, or electrons, must act gravitationally as if each were independent of its neighbours. According to our present point of view as to the constitution of matter, the known

smallness of the electrons compared with the distances separating them enables us to infer the accuracy of the law at all sensible distances directly from its astronomical exactness.

Gravitational forces are extremely minute in comparison with electrical forces, and Sir Oliver Lodge shows (*Electrons*, p. 210) that the force between two electrons in contact, assumed for simplicity to be spherical, is about 10^{42} times as great as their gravitational force, that is to say, the ratio is about that of a million billion trillions to unity. In aggregates composed of electrons of opposite sign in equal numbers, these will neutralise each other's electrical effects at a distance, while their gravitative potential will be proportional to the sum of the squares of their charges, and therefore, as he points out, two bodies each consisting of 10^{21} electrons, that is to say, a thousand million billions, will attract each other, at any distance between the centres of gravity, with a gravitational force equal to that between two single electrons at the same distance. Lodge has expressed the enormous amount of the electric forces as compared with the gravitative ones in a form more easily grasped in the statement that: "If the opposite electricities were extracted from a milligramme of water and given to two spheres one mile apart, those two spheres would attract each other with a force equal to the weight of twelve tons."

We see, therefore, that gravitative effects must be of the second or a higher order of small quantities, as compared with electric actions. At first sight it would appear that these minute gravitational effects might be accounted for on the assumption that the simple expressions for the energy of the ether, from which its properties have been developed, are only to be regarded as approximations, but it is pointed out by Larmor that the introduction of higher order terms in the equations would involve the consequence that a train of radiation from a distant star would change its form and undergo dispersion in its course across space. If any such dispersion existed it would have to be extremely minute, since heavenly bodies emerging from eclipse show no changes of colour, as would be the case if there were sensible dispersion. Still, the ratio of the gravitational to the electric forces is so excessively small that the very minute dispersion which might exist without being observed might quite well be sufficient to account for gravitation. There is, however, as Larmor observes, another and fatal objection to accounting for gravitation on these lines, in that any action due to such higher terms would have relations with radiation, including a velocity of transmission of the same order as that of light, which is absolutely inadmissible.

If gravitational attraction is an action of radiational type

transmitted by means of the ether at a finite speed, it must be subject to aberration, just as in the case of light. Now it has been shown by Laplace (*Mécanique Céleste*, livre 10, § 22) that if a body of mass m_2 is moving with a velocity v relatively to a body of mass m_1 at a distance r, a wave of gravitational force emitted from the body m_1 with a velocity u, in addition to exerting a force $G\frac{m_1 m_2}{r^2}$ upon m_2 in the direction of the straight line drawn from m_2 to m_1, will give rise to a small force proportional to $\frac{m_1 m_2}{r^2}\frac{v}{u}$ acting in the direction opposite to that of v, and therefore tending to decrease the velocity v. If this is applied to the relative motion of the moon and earth, the result will be a continuous or secular decrease in the mean radius of the moon's orbit, giving rise to a secular acceleration in its orbital velocity. Such a secular acceleration was known from observation to exist, but if the whole of it were to be attributed to gravitational aberration, it would still be necessary for the speed of propagation of gravitational action to be at least seven million times as great as the observed speed of light. The amount of this secular acceleration, though small, was known to a very considerable degree of exactness, the results of comparatively modern observations being confirmed by records of lunar eclipses extending as far back as the year 729 B.C., the date of the earliest Babylonian record. Laplace's investigation of the motion of the moon relative to the earth appeared to account for practically the whole of this secular acceleration by the disturbing effects of the nearer planets on the moon's motion, and he therefore assigned the higher limit of one hundred million times the speed of light as the minimum speed of propagation of gravitational action. The argument for this higher limit has, however, been largely vitiated by the comparatively recent correction of Laplace's investigation by J. C. Adams, who shared with Leverrier the honour of demonstrating the existence of the planet Neptune from the deviations of Uranus from the orbit calculated for it, after allowing for the perturbing effects of the previously known planets. Adams found that a term in the approximation which Laplace had considered negligibly small led, when the approximation was carried to a further stage, to a sensible term, owing to the small neglected term having to be divided by another term, in itself negligibly small, and so giving rise to a quotient of sensible amount. Adams' laborious calculations showed that planetary perturbations accounted for little more than half the observed secular acceleration, and the remainder remained unaccounted for until Lord Kelvin pointed out that the tidal action of the moon on the earth

would cause a slow diminution in the earth's speed of rotation, and so give rise to an apparent increase in the moon's secular acceleration. It does not appear possible to calculate directly the effect of this tidal action, but even if it were neglected entirely, the speed of propagation of gravitation could not be much less than twenty million times that of light.

The question has been more recently discussed by R. Lehmann-Filhès (*Astronomische Nachrichten*, 1885) and J. von Hepperger (*Wiener Berichte*, 1888). The former, from consideration of the effect on the mean radii of planetary orbits, concludes that gravitational attraction cannot be propagated with a velocity less than five hundred times that of light; the latter, from a consideration of the effect on the mean radius of the moon's orbit, arrives at a million times the velocity of light as the minimum speed of propagation of gravitation. The existence of a finite speed of propagation would not, it may be noted, be of any assistance in accounting for the discrepancies referred to on page 418, if Newton's law is maintained in its original simplicity, in which the masses of moving bodies are assumed to have constant values unaffected by their motions. It has been shown, however, that the most noteworthy of these discrepancies may be accounted for by means of a generalised law of gravitational action transmitted with a finite speed, according to which the attraction between two heavenly bodies depends on their relative motion at each instant, as well as on the instantaneous distance between them, and on the values of their masses, assumed to remain constant (see Appendix N). These considerations acquire great interest and importance in view of the fact that the electric theory of matter leads to the conclusion that the inertia of a body is not an absolutely constant quantity, but that the electromagnetic mass of a body depends partly on its velocity; and, moreover, that the inertia in virtue of which it resists change of motion is a vector quantity depending on the direction of its motion relative to that of the force tending to change it, as well as on the numerical value of its velocity. They appear, in fact, to suggest the possibility that these discrepancies may ultimately find an explanation in this variability.

It has been very generally assumed that Laplace's argument precludes the possibility of gravitational action, whatever may be its nature, being propagated at a speed comparable with that of light. It is argued, on the other hand, by W. Wien,[1] that the speed of propagation could be observed experimentally only by

[1] *Amsterdam Proceedings*, 1900, p. 559; or *Ions, Electrons, Corpuscles*, Abraham et Langevin, p. 1065.

observing the resulting perturbations, if the intensity of gravitation increased and diminished instead of remaining constant. As the intensity actually remains constant, he maintains that the effect of a finite speed of propagation can only affect the perturbations due to the motions of the attracting bodies.

It appears to me that what Laplace's argument proves is the incompatibility with the results of astronomical observation of the existence, to any sensible extent, of aberration of gravitational attraction, and the same remark applies to the more recent investigations of R. Lehmann-Filhès and J. von Hepperger. It would, therefore, apply to any theory of gravitation which ascribed it to action of the nature of radiation, but would not apply to action, although transmitted at a finite speed through the ether, which would not give rise to aberration.

A course that naturally suggests itself is to look for the origin of gravitation in compressional ether waves, as, although this would be an action of the nature of radiation, the speed of transmission might then quite possibly be of the order required, if the compressibility were sufficiently minute. Larmor therefore tried the effect of introducing a compressional term into the expression for the potential energy of the ether, but found that this would lead to a force of repulsion between two material bodies, and of a different form numerically from the force of gravitation. Another assumption that might be made would be that there is a slight difference in amount between the charges of the positive and negative electrons in a molecule, but this again would lead to a repulsive force, as would any static effect, represented by a term in the expression for the potential energy, for the approach of two atoms will, as Larmor points out, in that case increase the strain, and therefore the stress, and hence would increase in a higher ratio the energy of strain which depends on the product of the strain and the stress. The mutual forces exerted by the atoms would therefore oppose their approach.

Lorentz has attempted in two different ways (*Amsterdam Proceedings*, April 1900) to arrive at an electrical explanation of gravitation on the basis of the electronic constitution of matter.

In the first of these he assumed gravitational attraction to be produced by electromagnetic waves of short length compared with the smallest distances for which the validity of the Newtonian law is to be assumed. This theory he abandoned, owing to its only accounting for gravitational attraction by the aid of assumptions involving a continuous absorption of energy by material bodies. I may point out that it would also involve, not only a rate of propagation corresponding to that of ordinary electro-

magnetic radiation, but aberrational effects of the same order as in the case of light.

In his other attempt Lorentz has recourse to a theory of earlier date due to O. E. Mossotti and F. Zöllner. The former, in a volume *Sur les force qui régissent la constitution intérieure des corps*, published in 1836, proposed to account for gravitation on the assumption that the ether, as well as matter, was of atomic constitution. The material atoms were assumed to repel one another according to the law of the inverse square, and likewise the ethereal atoms; but an attraction of slightly greater numerical value than either of these repulsions, also following the Newtonian law, was assumed between the ethereal atoms and the material ones. This hypothesis was simplified in 1882 by Zöllner in a volume entitled *Erklärung der universellen Gravitation aus den statischen Wirkungen der Elektrizität.* Each gravitating molecule was assumed to be composed of portions carrying positive and negative electric charges, and the attraction between a pair of unlike charges was assumed to slightly exceed in numerical value the repulsion between a pair of like charges. This hypothesis is applied by Lorentz to matter assumed to be built up of positive and negative electrons, and involves the identity of the rate of propagation of gravitational action with that of light, both in the free ether and in material media. It would not appear to necessarily involve aberrational effects, but it is difficult to see how other inadmissible consequences are to be avoided. For example, the gravitational attraction between two bodies should vary with the nature of the intervening medium to an extent which would be easily observable. Moreover, the hypothesis is not a development of the electromagnetic scheme, but an assumption imported into it in order to account for specific phenomena, a procedure equivalent, as it seems to me, to a complete abandonment of the definiteness and simplicity of the electron theory.

A remarkable consequence of assuming gravitation to be an action transmitted by the ether was pointed out by Maxwell in a note to his paper on "A Dynamical Theory of the Electromagnetic Field" (*Phil. Trans.*, 1865). He points out that the lines of force in a gravitational field containing two particles of masses M_1 and M_2 would be of the same form as those of a magnetic field containing two similarly placed magnetic poles of strengths m_1 and m_2, numerically equal to M_1 and M_2, the force being attractive in the former case and repulsive in the latter (see Appendix N). Let W and W′ be the intrinsic (potential) energies of the respective fields, X and X′ the respective forces acting during a displacement δx, and R and R′ the resultant forces at

corresponding points of the two fields. Then the work done in the displacements will be

$$X\delta x = \delta E, \; X'\delta x = \delta E'.$$

Now $\qquad X = -X'$, so that $\delta E = -\delta E'$.

Hence, where C is a constant,

$$E = C - E' = C - \sum \frac{H^2}{8\pi}\delta\tau,$$

where $\delta\tau$ is the element of volume, and the summation is extended throughout the magnetic field.

Now $\qquad R = -R'$, so that $H^2 = R'^2 = R^2$,
and therefore

$$E = C - \sum \frac{R^2}{8\pi}\delta\tau.$$

The intrinsic energy of a gravitating field must therefore be diminished by the existence of a resultant gravitating force. But the intrinsic energy is an essentially positive quantity, and therefore it follows that those parts of space where the resultant gravitational force is null, as in the equilibrium position between two attracting bodies, must possess an intrinsic energy per unit volume greater than $R^2/8\pi$, where R is the greatest possible value of gravitational force in any part of the universe. The assumption, therefore, that gravitation arises from an action transmitted by the ether, necessarily leads to the conclusion that every portion of this medium possesses, when undisturbed, this enormous intrinsic energy, and that the presence of material bodies so influences the medium as to diminish this energy whenever there is resultant gravitational attraction. Maxwell was unable to understand in what way a medium could possess such properties, and was therefore led to abandon the idea of seeking for an explanation of gravitation in action transmitted by the ether.

Now the electrical theory of matter accounts for the second property, since the presence of electrons gives rise, as we saw in Chapter IX., to a decrease in the elasticity of the ether within the field of their influence, and this, as we have seen, necessarily involves a corresponding decrease in the intrinsic energy.

According to Sir Oliver Lodge's estimate, the value of the constant C, the intrinsic energy of the ether when in the undisturbed condition, is of the order of 10^{33} ergs per cubic centimetre, and it appears far more likely that this estimate is too low than too high. The highest value of the gravitational

force R actually known to us is its value at the surface of the sun, where it is of the order of $2 \cdot 5 \times 10^8$ C.G.S. units, or about twenty-seven times as great as at the earth's surface. The value $R = 3 \times 10^3$ or $R^2 = 9 \times 10^6$ would therefore give a minimum value for C. There is, however, no reason to suppose that the sun is the largest gravitating mass in the universe, so that the probable minimum value allowable for C must be greater than this. Now, if $C = 10^{33}$, R^2 may exceed 9×10^{32}, or R may exceed 3×10^{16}. That is to say, R may be as much as 10^{10}, or ten thousand million times as great as the gravitational force at the sun's surface. It is to be noted, moreover, that the conclusion indicated by electrical theory is not that 10^{33} is the most probable value of Maxwell's constant C. This value is indicated as the smallest one that can be ascribed to C, which may actually be very much larger.

Under these circumstances it would hardly appear that the amount of intrinsic energy which must be possessed by the ether, if it is the medium for the transmission of gravitational attraction, could by any possibility become a valid objection. It is, however, of interest to inquire whether it is possible to assign any definite superior limit to gravitational force in the universe. Now, if the number of bodies in our ethereal universe is indefinitely great, the problem of determining the value of R at any given point of space involves the determination of the action at a given point of the effect of an infinite number of finite attracting masses. It was first pointed out by C. Neumann[1] that the solution of this problem in accordance with the Newtonian law may become indeterminate. The problem has been very fully investigated by H. Seeliger,[2] who shows that if the total quantity of matter in the universe is infinite, the force of gravitation may not merely become indeterminate at certain points, but may attain infinite values.[3] He therefore attempted, but without success, to find some modification of Newton's law which would obviate this difficulty without coming into conflict with the results of astronomical observation. It has been pointed out by Föppl[4] that the difficulty may be overcome by postulating the existence of negative as well as positive masses in the universe. If the sum of such negative masses were assumed to be numerically equal to the sum of the positive masses, then the gravitational lines of force, like the lines

[1] *Leipsic Abhandlungen*, 1874.

[2] *Astronomische Nachrichten*, 1895; *Münchener Berichte*, 1896.

[3] A simple proof of this is given in Appendix O. It leads to the conclusion "that the mean density of ponderable matter throughout any very large spherical volume of space is smaller, the greater the radius; and is infinitely small for an infinitely great radius." (Lord Kelvin, *Baltimore Lectures*, p. 267.)

[4] *Münchener Berichte*, 1897.

of electric force, would all start from positive masses and terminate in negative ones. We shall see that the possibility of the existence of both positive and negative gravitating masses is a natural consequence of some of the most promising of the theories by which it has been attempted to account for the phenomenon of gravitation. Moreover, it does not give rise to any difficulties, for negative masses would attract each other according to the same law as positive masses, and in the present stage of our knowledge it would be impossible to distinguish a system composed of negative masses from one composed of positive masses. The presence of visible positive and negative masses near together would be immediately observable, since masses of opposite sign would repel instead of attracting each other; but a probable, if not a necessary, consequence of this repulsion would be that systems of opposite sign would become separated by distances too great for their mutual action to be observable.

It is pointed out by A. Schuster (*Nature*, 1898) that the existence of negative mass would afford an explanation of the origin of the rotational velocity of the solar and other stellar systems. Unless we assume that these rotational velocities were directly impressed by creative power, their origin, in accordance with dynamical principles, must have been accompanied by the generation of equal and opposite moments of momenta in other systems, and in this case the question arises as to how the corresponding systems have parted company, if the gravitational forces in the universe are entirely attractive.

Lord Kelvin[1] has investigated the consequences of supposing matter (positive only), in the form of luminous and extinct stars, amounting to a thousand million times the Sun's mass to be included within a sphere with the earth as centre and a radius of $3{\cdot}09 \times 10^{16}$ kilometres, which would include all stars having an annual parallax of not less than one one-thousandth of a second of arc, that is to say, all the visible stars down to the 16th magnitude. Assuming for simplicity the matter to consist of a thousand million stars of equal masses, he finds that if they had been at rest thousands of millions of years ago, at distances from one another very great in comparison with the radius of the sphere now including them, and so distributed that they should now temporarily be equally spaced throughout this sphere, the result of the gravitational interaction would be to give them a mean velocity (reckoned as the square root of the mean of the squares of their actual velocities) of 50·4 kilometres per second.[2]

[1] *Baltimore Lectures*, p. 268 and Appendix D, or "On the Clustering of Gravitational Matter in any part of the Universe," *Phil. Mag.*, Aug. 1901.
[2] See Appendix O.

Lord Kelvin observes that this is not very unlike what we know of the stars visible to us. Moreover, a non-uniform distribution would give rise to greater velocities than would the uniform distribution. He therefore arrives at the conclusion that, however much matter there may be outside the sphere of $3{\cdot}09 \times 10^{16}$ kilometres radius, the total quantity within it must lie between 100 million times and 2000 million times the Sun's mass.

It was an old and celebrated hypothesis that if we could see far enough into space the whole sky would appear to be covered with the discs of stars, and that the reason that this was not so was to be sought in an absorption of the light of the more distant stars in its passage through the ether. As we now know that no such absorption takes place, it has been suggested that the limitation in the number visible is a proof of the finite extent of our universe. This is, however, an incorrect conclusion.

Let us suppose that one thousand million stars are all luminous, and of the same size and brightness as the Sun, and further that they are uniformly distributed throughout the sphere of radius $3{\cdot}09 \times 10^{16}$ kilometres, and that there are no stars outside this sphere, and let us on these suppositions find what would be the total amount of starlight in comparison with sunlight. Let n be the number of stars per unit of volume, the radius of each being a. The number in a shell of radius q and thickness dq will then be $n{\cdot}4\pi q^2 dq$, and the sum of their apparent areas as seen from the centre will be

$$\frac{\pi a^2}{q^2} n{\cdot}4\pi q^2 dq = n{\cdot}4\pi^2 a^2 dq.$$

Now

$$n{\cdot}4\pi^2 a^2 \int_0^r dq = n{\cdot}4\pi^2 a^2 r.$$

Again, if N be the total number in the sphere of radius r

$$n = \frac{N}{\frac{4\pi}{3} r^3},$$

so that

$$n{\cdot}4\pi^2 a^2 r = N{\cdot}3\pi \frac{a^2}{r^2}.$$

If a be the ratio of the sum of the apparent areas of all the stars to 4π,

$$a = \frac{3N}{4} \frac{a^2}{r^2},$$

and $(1-a)/a$, which is very approximately equal to $1/a$, is the ratio of the apparent area not occupied by stars to the sum of the apparent areas of their discs. Therefore a is the ratio of the apparent brightness of the starlit sky to the brightness of the Sun's disc, and Lord Kelvin points out that, with the assumed values of r and a, cases of one star eclipsing another wholly or partially would be so rare as to be negligible, and moreover, this negligible deduction would be almost entirely annulled by diffraction, which makes the total light from two stars, one of which is eclipsed by the other, very nearly the same as if the more distant one were seen clear of the nearer one.

Now $N = 10^9$, and $a = 7 \times 10^5$ kilometres, so that $r/a = 4{\cdot}4 \times 10^{10}$ and therefore $a = 3{\cdot}87 \times 10^{-13}$. But if r is varied while the density of the matter remains the same, N varies as the cube of r, and therefore a varies simply as r, so that in order to make a as great as $3{\cdot}87/100$, or, say, the sum of the apparent areas of discs 4 per cent. of the whole sky, the radius must be $10^{11}r$, or $3{\cdot}09 \times 10^{27}$ kilometres. Now light travels 300,000 kilometres per second, or $9{\cdot}45 \times 10^{12}$ kilometres per year. It would therefore take $3{\cdot}27 \times 10^{14}$ years to travel from the most distant stars within the sphere to the centre. Even if we attributed to each star the extremely outside estimate of a life of a billion years as a luminary, this would be more than three hundred times the life of the star. Even, therefore, if we supposed all the stars to have become luminous at the same time, more than three hundred times the life of a star would pass before light would begin to reach the earth from the outlying stars, and at no one instant would light be reaching the earth from more than a very small proportion of all the stars.

Since gravitation does not appear likely to be capable of inclusion in the electromagnetic scheme, it becomes necessary to consider the more promising of the theories propounded independently of that scheme, and to attempt to correlate them with it.

It will, however, be advisable, for reasons which will presently appear, to show that the phenomena of cohesion may be accounted for as a direct consequence of gravitational attraction, quite independently of any particular theory of the latter. It was shown by Lord Kelvin in a "Note on Gravity and Cohesion," communicated to the Royal Society of Edinburgh in 1862, that cohesion will be a necessary consequence of gravitational attraction, provided only that the space occupied by the atoms of a material body is sufficiently small in comparison with its bulk; that is to say, if the atoms are sufficiently small in comparison with their mutual distances. To prove this, he imagines two homogeneous cubes of matter to be placed with one side of each in perfect contact with one side of the other, while one-third of the contents

of each cube is condensed into a very large number, say n, of square bars perpendicular to the interface, and so placed that each bar of one group has an end in complete contact with an end of a bar of the other group, the remaining two-thirds of the matter of each cube being supposed to be, for the moment, removed. The mass of each bar will then be $1/3n$ of the original mass of either cube, and however small may be the masses of two such bars, the attraction between them, per unit of contact area, may be increased without limit by diminishing the sectional area of the two bars while keeping their masses constant. Now the total attraction between the two groups is greater than the sum of the attractions between the pair of bars, that is to say, greater than n times the attraction between one pair of conterminous bars. The whole attraction between the two cubes may therefore be made to attain any value, however great, by sufficiently diminishing the sectional areas of the bars while keeping their number and the mass of each constant.

Now suppose another third of the mass condensed into bars perpendicular to another face and to the first set, and the remaining third into bars perpendicular to both these sets, and therefore to the two remaining faces of the cube, and let this be done with each of the cubes. Then, if the cubes are packed together in the relative positions which they would naturally tend to assume, viz. so that the bars fit end to end, the cohesion force per unit of interface may be increased to any extent by sufficiently increasing the ratio of the space unoccupied within the boundary of each cube to the space occupied by the bars. This particular arrangement was selected by Lord Kelvin only for the sake of definiteness and simplicity, and, as he points out, the argument applies to any heterogeneous or discrete atomic arrangement of matter, provided the ratio of the space occupied by the material atoms to the space unoccupied by them (and therefore occupied by ether) is sufficiently great. According to our present view of the constitution of matter, and of its excessive flimsiness in comparison with the massiveness of the ether, the correlation thus established between the phenomena of cohesion and gravitation is of the most convincing and satisfactory nature.

It was pointed out in Chapter V. that Newton expressly postulated the necessity of some medium for the transmission of gravitational attraction, and this is very plainly expressed in his third letter to Bentley, dated 25th February 1692–3.

"It is inconceivable that inanimate brute matter should without the mediation of something else, which is not material, operate upon and affect other matter without material contact, as it must do if gravitation, in the sense of *Epicurus*, be essential and inherent in it.

And this is the reason why I desired you would not ascribe innate gravity to me. That gravity should be innate, inherent, and essential to matter, so that one body may act upon another at a distance through a *vacuum*, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it. Gravity must be caused by an agent acting constantly according to certain laws; but whether this agent be material or immaterial, I have left to the consideration of my readers."

In query 21, at the end of Newton's *Opticks*, the pressure of an ambient medium is definitely suggested as the origin of gravitation.

The cosmical views of Huygens (see Chapter IV.) naturally led him, on the publication of Newton's *Principia*, to seek the origin of gravitation in some kind of ethereal connection extending across the space separating the mutually attracting bodies. While admitting that Newton's law was in correspondence with the observed facts, as far as the attraction between the heavenly bodies or other masses which were far apart in comparison with their dimensions was concerned, he was unable to persuade himself that there was any likelihood of the ethereal connection being equivalent to so simple a law in the case of bodies near together, or of different parts of the same body. The gradual accumulation of evidence on the subject shows that the law as stated by Newton holds good for very much smaller distances than Huygens supposed, and that it leads to conclusions in accordance with observations, when applied to the constituent molecules even of a solid mass. This, however, does not imply its applicability to masses whose distances apart are comparable with their dimensions, and it does not appear at all probable that this is the case. The reason of its applicability to the constituent particles of material bodies is to be found, as pointed out at the beginning of the present chapter, in the fact that the distances between the ultimate particles of which matter is composed are very great in comparison with their dimensions. This extension of the law to every particle of matter in the universe, when stated by Newton, was a bold but unproved hypothesis. It now follows, as a consequence of its proved astronomical validity, coupled with the similarity in the order of the relation of dimensions to distance for the heavenly bodies and the sub-atoms of matter.

One of the earliest attempts to formulate a dynamical theory of gravitation was made in an elaborate essay by the elder John Bernoulli in 1734, and was based on a combination of Cartesian vortices with the impact of outside particles. Its chief interest,

other than the mathematical fame of the author, consists in the definite formulation of the hypothesis that matter is built up out of a "primal element perfectly uniform and without structure," and which offers no resistance to motion of matter, and is therefore a perfect fluid.

The "Essai d'une Nouvelle Physique Céleste" will be found in Bernoulli's collected works, and an abstract of its contents is given in a report on Kinetic Theories of Gravitation by W. B. Taylor (*Smithsonian Institute Reports*, 1876). This report contains an interesting account of Newton's further conjectural suggestions, and of the hypotheses (of very varying merit) propounded by various writers prior to that date. It may not, however, be amiss to warn the reader against accepting the statement of the conditions of the problem as authoritative. The criticisms and conclusions also call for discernment.

Before proceeding to the consideration of the recent non-electromagnetic, although ethereal, hypotheses which appear to me to afford the most promise of an ultimate solution of the problem, I must devote a short space to a theory which, in spite of some obvious deficiences, held the field for a couple of centuries as the only one which showed any promise of even an approach towards a solution of the problem of gravitation. This is Le Sage's theory of extramundane corpuscles, which was propounded in the year 1750, and was recently resuscitated for a time by Lord Kelvin. According to this hypothesis, gravitation is assumed to be due to the impact of extremely minute corpuscles flying through space in every possible direction. The impact of these particles upon an isolated heavenly body would give rise to an equal impetus in all directions, so that there would be no resultant force. In the case of two bodies, however, each would screen the other from a portion of this corpuscular bombardment, so that the bodies would be attracted to each other. As the impulse due to the impact is proportional to the effective area, the latter must be proportional to the mass. In order that this condition should be fulfilled, it was assumed that the molecules of material bodies are extremely small relatively to the distances separating them, so that only a very small proportion of the corpuscles would be arrested during the passage through a material body, and the number so arrested might then be taken to be proportional to the number of molecules in the body.

Two such bodies would arrest more of the particles moving in the direction of the line joining them than a single one, but of the particles traversing both in a given direction, the first would arrest a certain fixed proportion of the stream, while the second body, if of the same density, would arrest the same proportion of

the stream falling on it. The stream on the second would, however, be less than on the first, by the number stopped by the first, and therefore the number stopped by the second would be less, relatively to its dimensions, than the number stopped by the first. That is to say, one body would act as a screen to the other. The objections to this theory have been pointed out by Lord Kelvin, who, however, resuscitated the theory for a time by showing how it might be so modified as to obviate some of them. The particles would have to move at enormous speeds, as otherwise they would act as a resisting medium and impede the planetary motions to an observable extent.

This point has been very fully investigated by Rysánek (*Repertorium für experimental Physik*, 1887) on the basis of the kinetic theory of gases. He assumes the speeds of the corpuscles to be distributed in accordance with Maxwell's law, and arrives at the conclusion, based on the accordance between the observed and calculated orbit of Neptune, that if the density of the particles were that of the ether, their mean velocity could not be less than 5×10^{19} centimetres a second; that is to say, it would have to be considerably more than a thousand million times the velocity of light. This would require an enormous and continuous supply of energy from some agency outside the universe, and if, as Maxwell has pointed out, the rate of supply of corpuscles were to vary for any reason, the value of every force in the universe would suffer change. The preservation of the universe would therefore be effected only by the unceasing expenditure of enormous quantities of work, so that the conservation of energy would become apparent only, and be simply a sort of movable equilibrium between supply and destruction.

Lord Kelvin pointed out that if the particles were perfectly elastic there would be no screening, because in that case a body would reflect as many as it stopped. If they are inelastic, as Le Sage assumed, what becomes of them after impact, and why are not bodies always increasing in size by their accumulation? Moreover, if inelastic or only partially elastic, the development of heat would be so enormous that all the matter in the universe would be volatilised in the course of a few minutes.

Lord Kelvin's modification consisted in replacing the small solid particles by vortex atoms carrying away with them the energy generated by impact, in the form of an increased intensity of vibration, of an amount sufficient to compensate for the loss of translational energy. He also suggested that this acquired vibrational energy might be again transformed into translational energy by interactions in infinite space.

The details of Le Sage's theory have since been very thoroughly

investigated by Sir George Darwin in a paper on "The Analogy between Le Sage's Theory of Gravitation and the Repulsion of Light" (*Roy. Soc. Proc.*, 1905). He has considered the case of two spheres subjected to the bombardment, evaluating separately the effects of the normal and tangential components of the several impacts. The conclusion arrived at is that neither the normal nor the tangential components give rise to forces of attraction, between the spheres, which vary rigorously as the inverse square of the distance between their centres. The sum of the two effects will not in general ensure the rigorous truth of the law of the inverse square, nor will it ensure equality of action and reaction between the two spheres. The inequality of action and reaction will be obviated if the elementary portions of matter are rigorously of the same size. There is one limiting case in which both the law of the inverse square and that of action and reaction are rigorously fulfilled. This will be the case if the corpuscles are absolutely devoid of elasticity, and the roughness of the spheres be such as absolutely to annul the tangential velocities of the corpuscles on impact, that is to say, if the whole energy of momentum is absorbed in the impacts. This hypothesis would, however, leave in its acutest form the necessity for the continuous supply of external energy.

The fact that a hypothesis so artificial, and containing so many defects, as that of Le Sage should be deemed worthy, after the lapse of more than two hundred years, of serious investigation by such eminent physicists, will serve to impress the reader's mind with the extreme difficulty presented by the problem of forming a dynamical representation of gravitational action. Many other attempts, in addition to those already referred to, have been made to improve or modify Le Sage's hypothesis. One of the most elaborate of these was the hypothesis propounded by Jarolimek (*Wiener Berichte*, 1883), according to which the gravific corpuscles were assumed to be identical with the ultimate sub-atoms of matter, and the Newtonian law of gravitation was deduced from the extremely artificial assumption that the distribution of the velocities of the corpuscles was such that every portion of space would contain n^2 times as many corpuscles having an effective free path of length r, as the number having an effective free path of length nr. In addition to this artificiality the hypothesis only removes some of the defects of that of Le Sage, and is even less reconcilable with present views of the constitution of matter.

There remain for consideration two hypotheses, both of which have one feature in common with that of gravific corpuscles, in that they both appear necessarily to involve the assumption

that our ethereal universe only forms a portion of a greater universe, and to compel us to contemplate definite physical actions extending beyond the extremest limits of that ether which fills all that we know as space.

A few years ago such hypotheses would have been regarded as speculations too wild and improbable for the pages of a romance. But so would a prediction that it would ever be possible to obtain experimental evidence of an effect so minute as a shortening of the earth's diameter of eight thousand miles by an amount of about six and a half centimetres in the direction of its orbital motion at each instant, or that we should one day come to regard every smallest speck of matter as very like a star system in miniature, with much the same kind of changes and convulsions occurring within it as those which we observe on a larger scale in the birth and decay of the solar systems of our starry universe. If these are accepted as legitimate theories—and they are regarded by those best entitled to express an opinion as the most complete and satisfactory representations of material phenomena attainable in the present state of our knowledge—there is no reason to recoil from contemplation of the possibility that our universe itself is but as a speck in a still greater one.

The hypotheses themselves are of a simple character, and although they are not comprehended within what I have referred to as the electromagnetic scheme, not only are they consistent with it, but either of them would appear to supply just the additional forces upon the electrons which would fix their sizes (see Appendix P) and so make the scale of material atoms determinate.

The first of these was propounded by Professor W. M. Hicks in the *Cambridge Philosophical Society's Proceedings* for 1880. It is founded on some investigations by Bjerknes, extending from 1863 to 1875, of the problem of the *pulsations*, as he called them, of two spheres in a fluid, pulsations being defined as consisting of periodic changes of volume, just as vibrations consist of periodic changes of position. Hicks succeeded in developing a method of rigorously determining the action of the two spheres, and showed how the action might be applied to explain the gravitation of matter consisting of Kelvin vortex atoms. It is therefore also applicable to matter built up of electrons, as the property possessed by two pulsating spheres in a fluid, of acting on each other with a force whose principal part varies inversely as the square of the distance, is common to all pulsating bodies. This will apply to slightly compressible, as well as to incompressible, fluids.

It must be assumed that the vortex atoms or electrons pulsate

with a constant period, and that none of the phases differ by more than 90° for any one system of mutually attracting bodies, or that if such differences once existed, they have been eliminated. A material system containing a mixture of atoms pulsating with every possible difference of phase would possess some very curious properties. For example, although two out of three atoms might attract the third, they would not necessarily attract each other. According to this theory of gravitation, it would therefore be quite possible for the bodies of any one star system to attract bodies belonging to one set of systems, repel those belonging to another, and exert no influence on those belonging to a third, and yet at the same time the bodies of each of the systems would obey the ordinary law of gravitation within the limits of their own system. The theory therefore provides for the existence of negative, as well as positive, gravitational masses, and also involves the possibility of one gravitating mass being neutral with regard to another. If we call positive masses those which would attract and be attracted by the bodies of our solar system, it would not be likely that any negative masses could be driven towards our system with sufficient energy of motion to force their way within it against the repulsion to which they would be subject. It would, however, be quite possible for masses which would be neutral relatively to the bodies of our system, driven by repulsion from distant portions of the universe, to find their way to our vicinity. That is to say, the theory suggests the possible existence of weightless matter within our system.

In order that the ordinary law of gravitation may be obeyed within the universe as known to us, the phases of the pulsations must, as stated above, be restricted to differences less than 90°, and Sir Joseph Larmor suggests that—

"we may imagine (ideally) the pulsation to have been applied initially over the outside boundary of the ethereal universe, and thence instantaneously communicated throughout the incompressible medium to the only places that can respond to it, the vacuous nuclei of the electrons; and we can even imagine the pulsations thus established as spontaneously keeping time and phase ever after, when the exciting cause which established this harmony has been discontinued."[1]

On the assumption that magnetic force is to be identified with twistless ether flow, the radially vibrating field would, as far as this twistless backward and forward pulsatory flow is concerned, resemble a magnetic field of alternating type in which the electron would correspond to a unipolar magnet. It would, however, differ

[1] *Phil. Trans.*, vol. cxc., A, 1897, p. 217.

from an alternating magnetic field in the very important respect that it would not be accompanied by alternating electric displacement, and would not therefore give rise to any magnetic effects. Larmor suggested that the unsatisfactory feature of this radial quasi-magnetic field was that it was introduced for the sake of gravitation alone, and that it did not present itself in any direct correlation with other physical agencies. This objection has now been removed by the fact that the action on the electrons of forces of other than electromagnetic origin appears to be requisite in order to make their scale determinate, which would be effected by the forces arising in this manner.

It was subsequently shown by A. H. Leahy[1] that the attraction between two spheres pulsating in the same phase and with equal periods changes sign, and becomes changed into repulsion, when the distance between them exceeds half a wave-length. It would therefore be necessary for the waves to be so long that the half wave-length should exceed the greatest distance for which Newton's law of gravitational attraction is known by astronomical observation to hold good.

An attempt was made by Professor Challis, in a work published in 1873, to develop a theory of gravitation on the basis of compressional waves maintained from a source outside the universe, and setting up vibrations in matter assumed to consist of hard spherical atoms endowed with inertia and immersed in a perfectly fluid ether also possessing inertia. Mathematical difficulties, however, prevented him from deriving any definite demonstrable results from these assumptions.

Dr C. V. Burton (*Phil. Mag.*, Feb. 1909) has shown that the establishment and maintenance of the pulsations of the electrons within the required phase limits may be accounted for on the assumption that periodic compressional waves are continuously traversing the ether of our universe, entering from, and passing on into, the unknown beyond its limits. The ether is assumed to be very slightly compressible, the compressibility being assumed to be somewhat increased by the presence of gravitating matter. This additional compressibility is obtained by assuming the ether within the nuclei of the electrons to be of slightly less density than in space free from matter. In addition to providing the additional compressibility required for the excitation of pulsations by the variations in ether pressure due to the alternate waves of compression and rarefaction, Dr Burton considers that this assumption makes the free mobility of electrons through the ether more easily conceivable than on the assumption of vacuous

[1] *Cambridge Phil. Trans.*, vol. xiv., 1885, pp. 45 and 188.

nuclei. It also has the advantage of greatly simplifying the dynamical conditions of the problem.

If, within a region bounded by any closed surface, a mass m of matter could be created, a certain amount of ether, which would be proportional to m, and might be called Fm, would be extruded from the region. The quantity F is defined as the extrusion of ether per unit mass of matter.

The compressibility of the ether is assumed to be so small that the slight variations in density will not give rise, to any sensible extent, to electromagnetic waves. The whole motion will therefore be twistless. Now let us suppose in the first place that the extrusion F is null, so that a given volume occupied by matter contains the same amount of ether as would be the case in the absence of the matter.

The ether affected by the changes in pressure due to the passage of the waves will then, as it surges backwards and forwards, merely carry with it any matter that may be present. There will be no tendency for the distributions of strain, or strain forms, constituting the electrons either to outrun or lag behind the bodily excursions of the ether, so that the wave motion will have no tendency to accelerate the electrons relatively to the ether, and therefore every portion of matter will continue in its state of rest, or of uniform or accelerated motion relatively to the ether, exactly as if no wave motion were taking place. It appears, therefore (at least in the case of sufficiently great wave-lengths), that the passage of such waves could not be detected in any conceivable manner, since they can give rise to no phenomena perceptible to our senses.

We will now suppose the extrusion F to have a finite value. The mean density of the ether in a region containing matter will then differ from the density of free ether, and therefore a general bodily acceleration of the ether will tend to produce an acceleration of the material molecules with respect to the ether.

Burton points out that, since the compressibility is assumed to be extremely small, the proportional changes in the pressure and in the constitutional energy of the medium will be great, relatively to the proportional changes in density, so that if the amplitude of the waves were not extremely minute, relatively to the speed of transmission, effects might result which would be palpable to observation. This is, however, obviated by the immense length which has to be assumed for the waves in order that all heavenly bodies within sensible reach of each other's gravitational action should have their electrons pulsating in corresponding phases, the result of this immense length being

that, even throughout an astronomically considerable region, the ethereal pressure, and therefore the speed of electromagnetic radiation, may be treated as uniform at each instant.

Having thus arrived at the conclusion that these waves will give rise to no observable effects of any kind, we have to consider the secondary pulsatory effect resulting in gravitative action.

On the assumption made that the whole space considered is so circumscribed that its dimensions are small in comparison with the length of the waves, the primary pressure variation will be sensibly in the same phase throughout. It is then shown that when the distance between two bodies is small compared with the wave-length, the attraction between them will only deviate from the law of the inverse square to the second order of small quantities. If this condition were not realised, and assuming the waves to be travelling indifferently in all directions, it is found that there would be a gradual extinction of gravitational attraction as the distances between the bodies increased to a sufficient extent to become comparable to the wave-length.

The only possible way of testing the theory by means of numerical results is to start with the constants whose values are known more or less approximately, and then to proceed, at first mainly by conjecture, and then by trial and error, and see if the various quantities concerned can be so determined as to give results which are not demonstrably false.

To start with, there is Cavendish's constant of gravitation G, defined by the relation

$$G = \frac{Fl^2}{mm'},$$

where F is the measured force of attraction between two bodies of masses m and m' separated by a distance l, its value being $6 \cdot 66 \times 10^{-8}$ C.G.S. units. The density of the ether is assumed to be 10^{12}, and its constitutional energy 10^{33} (see p. 186), and the wave-length is taken to be $1 \cdot 5 \times 10^{17}$ C.G.S. or ten thousand astronomical units, the astronomical unit being the radius of the earth's orbit, about ninety-three million miles. The length of a wave is therefore taken to be nearly a billion miles.

It will be of interest to note the values which have to be given to some of the more important quantities in order to fulfil the requisite conditions.

The velocity of a compressional wave in ether is taken as $2 \cdot 19 \times 10^{27}$ C.G.S. That is to say, a light wave travelling at a speed of about 186,000 miles a second would take more than two thousand million years to travel the distance traversed by a

compression wave in a second. The frequency is taken to be $1{\cdot}29 \times 10^{16}$ C.G.S.

The bulk modulus of elasticity of the ether has the value $4{\cdot}78 \times 10^{66}$ C.G.S., or $4{\cdot}72 \times 10^{60}$ times the pressure of the atmosphere, and the maximum deviation of the ethereal pressure from its mean value is as great as $3{\cdot}9 \times 10^{45}$ C.G.S., or $3{\cdot}8 \times 10^{39}$ atmospheres. The additional ethereal compressibility per unit density of matter present is, however, only $4{\cdot}2 \times 10^{-66}$ C.G.S., and the maximum ether displacement during the wave motion is only ·69 micron.

The total energy per cubic centimetre of the primary disturbance has the value $1{\cdot}6 \times 10^{24}$ ergs, while the energy passing in a second across an area of a square centimetre at right angles to the direction of transmission has the enormous value of $3{\cdot}4 \times 10^{51}$ ergs, or $3{\cdot}4 \times 10^{41}$ kilowatts, an amount of energy exceeding by many million times the sun's entire store of available heat.

Dr Burton observes that:—

"We have become accustomed to the idea that our rapid motion through a medium of enormous density not only fails to provide us with a useful source of energy, but defies the most refined attempts to detect it by means of terrestrial observations. And there is nothing more inherently improbable in the notion that our universe may be traversed by waves of enormous energy, perceptible to us only by means of a minute secondary effect—gravitation. That we should be unable, even in our dealings with this feeble residual phenomenon, to extract with continual profit the minutest portion of the energy so abundantly propagated, is a view which may appear to us somewhat ironical, but which is not out of harmony with the trend of modern physical conceptions."

Provided the forces exerted in a field of gravitation act in the same direction on the positive and negative electrons, and are of magnitudes proportional, respectively, to the masses of such electrons, the identical acceleration of the two denominations will follow, without additional forces of electrostatic type being called into play. There will then be no dissipation of energy, and consequently no heating effect, as the result of the compressional waves. In fact, the only perceptible effect other than the gravitational one, would apparently be the addition of a small non-electromagnetic term to the expression for the inertia, or mass, of an electron. One of the effects of such a term would be to make gravitational mass, like inertia, dependent to some extent on the motions of the attracting bodies. As in the case of inertia, the effect would be extremely small, except for speeds approaching that of radiation. The considerations discussed in Appendix N

show, however, that small variations of this nature may very possibly account for some of the principal outstanding anomalies which are in apparent conflict with Newton's law.

K. Pearson[1] has shown that, by means of somewhat artificial and complicated assumptions, many optical and other electrical phenomena, as well as the phenomenon of gravitation, could be accounted for on the pulsation theory. The theory is elaborated with great ingenuity and mathematical power, but since the development of the far simpler and more comprehensive electromagnetic scheme known as the electron theory, it is rather of mathematical than of physical interest.

A pulsatory theory of gravitation would necessitate the extremely high rate of propagation previously indicated in order to obviate perceptible aberrational effects. This would, however, be provided for by the extremely small compressibility of the ether indicated by electromagnetic considerations. The wave-length shown by Leahy to be necessary would prevent sensible absorption in traversing any material bodies of the dimensions whose existence would appear to be possible within the regions of space accessible to observation of gravitational effects, so that no sensible variation in gravitational permeability could arise. The great wave-length and high speed of propagation would also prevent refraction or reflection of sensible amount.

The fluidity which must be ascribed to the ether suggests, as an alternative method of accounting for gravitation, the hypothesis of a steady hydrostatic pressure transmitted through it. The possibility that gravitation might in some way arise in consequence of such an action was vaguely suggested by Newton himself, and later by Young, and has formed the basis for many of the attempts which have been made to solve the problem. The foundation for a definite theory on these lines was provided by Lord Kelvin in a paper "On the Forces experienced by Solids immersed in a Moving Liquid" (*Roy. Soc. Edin. Proc.*, 1870), and was derived from his extensions of Helmholtz's investigations into vortex motion. Lord Kelvin shows that when a steady or "cyclic" twistless flow is once established through one or more apertures in a movable solid immersed in a liquid, it will continue unchanged for ever, however the solid be moved or bent, and whatever influences it experiences from other bodies, but that if there are two or more such solids within a finite region, mutual forces will arise in consequence of the flow. If such a

[1] *Quarterly Journal of Mathematics*, 1883; *Cambridge Phil. Trans.*, 1889; *Lond. Math. Soc. Proc.*, 1888; *American Journal of Math.*, 1898.

solid consisted, for example, of a very long narrow tube, the stream lines at its ends, that is to say, the lines of inflow and outflow, would be exactly similar to the lines of magnetic force due to the poles of a long bar magnet, the distribution of fluid velocity corresponding to the distribution of magnetic force. The velocity for distances great compared with the diameter of the tube, and small compared with its length, would therefore vary inversely as the square of the distance from the near end of the tube. Two such tubes would attract and repel each other according to the same law as two bar magnets, with the important distinction, however, that in the case of the magnets, like poles repel and unlike poles attract each other, while in the case of the tubes there would be attraction between like ends and repulsion between unlike ends (see Appendix P). A steady flow of ether, either into, or out of, every electron of a gravitating system of bodies would therefore give rise to an attraction between any two bodies of dimensions small compared with their distance, which would vary inversely as the square of this distance. Each electron would form a hydrodynamic "sink" or "source" of ether flow, and the attraction would be proportional to the product of the number of "sinks" or "sources" in one particle by the number in the other one, provided the "sink" or "source" were of the same magnitude for every electron. The attraction would therefore be proportional to the product of the numbers of electrons in the two particles, that is to say, to the product of their masses. It would therefore be in accordance with Newton's law of gravitation. The theory provides for the possible existence of negative masses, for a system whose electrons were sinks would repel a system whose electrons were sources. The possibility of two gravitating systems being neutral to each other would be excluded, for the only way in which a system could be neutral to a gravitating system would be by being composed of electrons which were neither sources nor sinks, in which case its constituent bodies would have no mutual gravitative action and would, as a further consequence, be without cohesion.

The difficulty of this hypothesis as a theory of gravitation consists in accounting for the existence of ether sources and sinks without the assumption of a continuous creation or destruction of ether within the electrons, and it is presumably for this reason that Lord Kelvin, in the paper referred to, put it forward rather as a hydrodynamic illustration of gravitational action than as a hypothesis by which gravitation might be accounted for. There are only two alternatives. We must either assume a slow, steady expansion or contraction of the electrons, or a flow of ether through every electron between our universe and a greater

universe of which it forms a portion. In the latter alternative, every electron, that is to say, every part of every body, in our universe must lie on the boundary, which means that the greater universe must be a four-dimensional space.

The hypothesis of expanding electrons has been worked out mathematically by G. H. Schott as an attempt to account for the existence of the forces, other than those of electromagnetic type, between the electrons, which appears to afford the simplest, if not the only feasible, method of explaining the definiteness of scale of the electrons indicated by the observed fineness of spectrum lines. His analysis appears to me to lead to the conclusion that either form of the source and sink theory, and also the pulsatory theory, is capable of accounting for this definiteness, and the question is dealt with in Appendix P. In his published paper Dr Schott adopts a rate of expansion which would lead to a doubling of the size of the electrons in about three hundred years, which appears too great to be admissible, and in a letter to me he suggests a rate which would double the size of an electron in three hundred million years. In order to obtain the observed value of the gravitation constant, it would be necessary to ascribe to the ether a density of more than 10^{100} times that of water. I am not aware of any reasons that could be adduced against so high a density. Lodge has shown good grounds for believing that it cannot be less than 10^{12}, or a billion times, the density of water, and that it is probably considerably higher. The assumption of expanding or contracting electrons would lead to a small variation of gravitational mass with the velocities of the attracting bodies, and therefore appears to offer as good a prospect as the pulsatory theory of a solution of the anomalies in planetary motions previously referred to. An indefinite expansion or contraction could hardly be considered a legitimate assumption, and Dr Schott assumes really a slow pulsation of extremely long period. This raises the question as to what would happen at the change from expansion to contraction, as in the case of so slow a pulsation there would necessarily be a stationary condition of neither expansion nor contraction lasting for a sensible period, during which gravitation, and consequently the cohesion of all material bodies, would be evanescent, which would apparently involve a catastrophic upheaval of the material universe. Such a slow expansion or contraction of the nuclei of the electrons might be supposed to arise from a steady slow decrease or increase of pressure applied to the ether over the whole boundary of the universe, a simple enough mathematical conception, but not one which it is easy to picture physically. Moreover, this would involve simultaneous expansion or contraction throughout our

universe, and would therefore not make provision for the existence of negative masses within it.

Riemann suggested in 1853 that gravitation might be explained by assuming a flow, through every particle of matter, from our universe into the unseen universe, "Aus der Körperwelt in die Geisteswelt" (*Gesammelte Werke*, p. 529), and Professor Schuster, in his communication to *Nature* respecting negative mass previously referred to, also suggested the second of the two alternatives, pointing out that it involves the existence of the fourth dimension, although he propounded it apologetically as "a holiday dream." I may therefore venture to include in this chapter some arguments in favour of this point of view which I had written out before becoming aware that two such authorities had already advocated it. Every theory, including Le Sage's, which offers any approach towards a dynamical explanation of gravitation, appears to have the common feature of involving what I may call the extra-physical conception of actions between our universe and a greater one within which it is contained. Now, if we are driven to such extra-physical concepts in order to account for gravitation, it appears to me that the nearest approach we can make to picturing a physical boundary to our universe is by means of the mathematical concept that the three-dimensional region constituting our material universe forms a boundary between two four-dimensional regions. It would then be related to these regions in the same manner that a surface in our ordinary three-dimensional space is related to the regions on either side of it, the surface, for example, of a heavenly body, which separates the interior region from external space.

Let us imagine this surface to be covered by a layer of ether modified in such a manner as to permit of the motion of material bodies in any direction along the surface of the heavenly body, but not in an upward or downward direction. Let us imagine further that all material bodies, including the inhabitants of this heavenly body, are of the same small thickness, and that the latter are only capable of being affected by sound, light, or heat waves travelling parallel to the surface. Such inhabitants would be able to travel freely over the surface, but would be unable to form any conception of their own thickness, or of thickness at all. They would live and move in two dimensions, and could not possibly form any physical conception of a third dimension. If, however, they were mathematicians, they might extend the geometry of their two-dimensional space and form a three-dimensional geometry, just as we can form a four-dimensional geometry. That is to say, they would be able to form a mathematical concept, though not a physical concept, of a three-

dimensional region. Now imagine a spring of water to rise from the ground at one part of this "flatland" and to disappear into the ground at another part. The water would then appear to the inhabitants to be created at its point of entry and to be annihilated at its point of exit.

Let us now imagine that we, and all other material bodies in our universe, are constrained by the relation of the electrons to the ether to move through the ether only in three dimensions, but that a fourth dimension exists into which we might move, except for such constraint. We could no more become aware of it, nor form a true physical conception of it, than could the inhabitants of "flatland" become aware of, or physically conceive, a third dimension. Let us suppose, however, that the ether itself is not subject to such constraint, but can flow out of or into the four-dimensional region, say through the nuclei of the electrons, these apertures being supposed equal for all positive electrons, and also for all negative electrons. It must be borne in mind that every point in our universe would, under such circumstances, be in contact with the four-dimensional regions separated by our three-dimensional universe, just as every point of the "flatland" we have imagined would be a point of separation between two three-dimensional regions. Let us further suppose that the flow is either outwards or inwards, respectively, at all parts of our universe near enough to us for gravitative interactions with our system to take place, and that at some very distant portion, beyond the reach of such action, the flow is inwards or outwards, respectively. The effect would be that all the bodies whose electrons were sources would attract each other according to the law of gravitation, and the same would apply to the distant bodies whose electrons were sinks. If, however, any bodies whose electrons were sinks were near enough to bodies whose electrons were sources, repulsion would take place according to the same law instead of attraction. The existence of negative masses is therefore provided for, and as the flow would be twistless, it could only be detected by its translational effects, such as accelerating or retarding the velocity of light travelling with or against the flow. The gravitational action is, however, so small and the density of the ether so great that any such effects would be far below the limits of our powers of observation.

At first sight it might appear that this form of the source and sink theory would not lead to the small variation of gravitational action with the speeds of the attracting bodies which appears to offer the most probable explanation of outstanding anomalies, but a little consideration will, I think, show that such a variation is to be expected. The nucleus of an electron, spherical when at rest in

the ether, becomes deformed when moving with a velocity v through the ether into an oblate spheroid of eccentricity v/c, while, as Searle has shown (*Phil. Mag.*, Oct. 1897), still retaining the property of acting at all external points as if its charge were concentrated at the centre. A change in the ether circulation through the electron would appear to be a necessary consequence of such deformation, and would lead to a variation of gravitational action with the speed.

CHAPTER XXIV.

THE PLACE OF MIND IN THE UNIVERSE.

THE physical theory which I have endeavoured to set forth in the preceding chapters has now been extended until it has been shown to be capable of correlating the whole complex of phenomena which constitutes the universe of what is commonly called inanimate matter. The development of this theory, considered from the standpoint now attained, presents two very distinctly marked stages. The first of these stages is represented by the mechanical scheme dealing with the relations of matter in bulk, and which may be considered as forming a first model, or representation, of the physical universe. Energy and matter, and the relations between them, form the subjects of this representation. The essential basis of this scheme was the central force system developed on the pattern of the Newtonian law of gravitation, and this being assumed, the scheme proved perfectly adequate to the correlation of the phenomena of matter in bulk. Galileo, Huygens, and Newton, all three minds of the very first rank, appear to have been the first among physicists to foresee the possibility of the development of such a mechanical scheme, working under definite established laws capable of being ascertained from the observed phenomena, and determining the correlation of natural phenomena without requiring the direct intervention of external intelligence. They were in consequence the first to effect the disentanglement of physics from theology by abandoning, in their writings, all attempts to deduce the solutions of physical problems from assumptions regarding the intentions of the Creator; but these writings make it perfectly clear that not one of them ever contemplated the possibility of dispensing with creative intelligence as antecedently necessary for the existence of such an ordered scheme. For example, Newton, in the *scholia* to the *Principia*, declares that the existence of a Being endowed with intelligence and wisdom is a necessary inference from a study of celestial mechanics, and that to treat of God is therefore a part of Natural Philosophy.

The perfection to which the mechanical scheme was developed in the course of the eighteenth century led to this antecedent necessity being lost sight of by many thinkers, and even by some minds of undoubted depth and acuteness. I am not, however, aware that this can be said of any physicist of the first rank.

The second stage is represented by the molecular scheme, which is concerned with the mechanism of transmission of the system of forces postulated in the first scheme. It is concerned with energy, ether and matter, and their mutual relations. We have seen that every portion of the first model, which once appeared to be ultimate and final, is itself of very complex construction, and that the uniformity which appeared to characterise the mechanical scheme is apparent and superficial, being a statistical uniformity only. We found that these several complexes can only be investigated by means of hypotheses developed by statistical mathematical methods, and leading to inferences respecting the relations between the mechanical parts of the first model. These relations can be verified or disproved by experience, and so provide the means of testing the validity of the various hypotheses. Those hypotheses which lead to conclusions in accordance with experience are provisionally accepted, but always subject to the possibility of their having to be rejected, should the results of their further development, in combination with those of the development of our means of observation, be found to lead to a conflict with experience.

The development of molecular theory thus involves the building up of a structure within a structure, the attempt to form models of every one of the parts of which the mechanical model is constructed. In the course of this procedure we may find various improvements suggested in our original mechanical model, and these improvements may entirely change its character. There is, however, one essential condition to which all such improvements must conform. The new mechanical model must remain mechanically consistent with itself, and must afford at least as complete and consistent a representation of purely mechanical phenomena as was given by the former one. Molecular theory must not be converted into a kind of sink for the disposal of mechanical phenomena for which it is difficult, or perhaps for the time impossible, to find a place in the mechanical model. Mechanical phenomena must be considered unexplained until they are mechanically accounted for. To push them out of sight into the underlying complex of molecular phenomena, and hope that they will right themselves, is entirely unscientific. It is equally unscientific to endeavour to explain apparent imperfections in the mechanical and molecular schemes by invoking the direct inter-

position of creative or directive intelligence. The whole course of scientific progress affords the most unequivocal evidence that the primal intelligence works by methods which are intelligible in the light of scientific investigation, or may be expected to become intelligible as our knowledge of nature extends, and our methods of inquiry undergo further development.

It was not to be expected that this should be recognised, as fully as at present, even by the great minds which first effected the disentanglement of physics from theology, and Newton himself fell into errors of the kind. In a general *scholium* at the end of the *Principia* he ascribes the initiation of the planetary motions to the direct intervention of the Deity, citing the great eccentricity of cometary orbits, and their journeyings into the depths of space, as arguments against their deriving their origin from ordinary dynamical actions. He also expresses the opinion that the cumulative effects of the observed planetary perturbations would ultimately lead to the destruction of the solar system unless they were specially adjusted by the direct intervention of the Creator.

The analysis subsequently developed by Laplace in the *Mécanique Céleste* proved conclusively that these perturbations were consistent with the dynamical stability of the solar system, and so obviated the necessity of resorting to the hypothesis of special intervention. In his later work, the *Exposition du Système du Monde*, Laplace laid the foundation for a dynamical theory of the origin of the planetary motions. He refers in the same work to the celebrated controversy between Newton and Leibnitz, in which the latter reproached the former for his narrow ideas of the wisdom and power of God, while Newton retorted by stigmatising Leibnitz's theory of *pre-established harmony* as a perpetual miracle. Laplace's comment on the dispute is that these various hypotheses had been rejected by posterity, while at the same time doing full justice to the mathematical investigations of both these great men. A copy of the *Système du Monde* was presented by Laplace to the great Napoleon, who appears to have been struck by Laplace's remarks on this controversy, and subsequently rallied him upon it, remarking, "Newton has spoken of God, but I do not find his name once mentioned in your book," to which Laplace replied that he had not required that hypothesis. The celebrated astronomer, M. Faye,[1] in recounting this anecdote, mentions that he had been informed by Arago that Laplace hearing, shortly before his death, that this anecdote was to be published in his biography, specially requested that it should be

[1] *Sur l'Origine du Monde*, chapter viii.

suppressed, foreseeing probably that his reply might be interpreted by the unthinking as referring to the existence of the Deity as an hypothesis only. Laplace was far too profound a thinker to entertain such a view.

Mechanical theory is not to be expected to come into touch at any point with intelligence, and therefore no indications of the existence of intelligence in the universe are to be anticipated in this first approximation, except as the necessary condition of its existence. The absence of such indications has sometimes been made the basis for an argument against the existence of creative and directive intelligence, and Haeckel, for example, has laboured this argument at considerable length. It is, however, perfectly obvious that its only validity is within the limits of frankly anthropomorphic conceptions. If the universe originated from a great but imperfect intelligence, it would be conceivable that continual intervention might be requisite for the purpose of providing for unforeseen contingencies, but the argument fails entirely if this intelligence is assumed to be perfect, and infinite.

Some thinkers have maintained that the concept of such an intelligence is so difficult, and so far removed from any phenomena within the limits of our experience, that it is simpler to assume that the universe, as we know it, has existed from all eternity, that it is itself the ultimate and self-maintaining reality. The untenability of this contention has already appeared in earlier chapters. It is absolutely irreconcilable with the principle of the dissipation of energy. This principle has been shown by observation to be universally valid in all physical experience, though Maxwell's "sorting demon" illustration has shown that it is not a necessary law of nature in an intelligently directed universe. It has been argued that the observed tendency which is formulated in the dissipation principle may possibly be counteracted by some other conditions present in the universe, but as yet unknown to us. The plausibility of this argument is at first sight increased by the indication on page 236 of purely molecular conditions which, although beyond our present experience, might conceivably tend to counteract the observed dissipation of energy. It is, however, shown in Appendix K that in any conceivable universe undirected by intelligence and containing molecularly constituted matter, at least one function, represented in our universe by the *entropy*, must necessarily tend to increase. That is to say, in our universe the principle of the dissipation of energy is, in the absence of intelligent direction, a practically *necessary* consequence of the molecular constitution of the matter which it contains. Our universe must therefore, in the absence of directive intelligence, have had a beginning and be tending towards an end, in the

same sense that this must hold good of a going clock actuated by a spring or a burning oil lamp. Our universe is not therefore, in the absence of intelligent direction, a conservative system, and since this is a practically necessary result of the molecular constitution of matter, it cannot be altered by any conceivable means, in the absence of intelligent direction.

I have employed the term *practically* necessary in reference to the mathematical demonstration that a tendency towards an increase in the entropy of the universe is a consequence of the molecular constitution of matter, because the dynamical investigation presents the actual increase of the entropy, not as a theoretical certainty, but as an infinite probability. It would contradict no law of nature if the entropy were to remain constant, or to diminish, except the actual observation that, as far as concerns the portion of the universe accessible to observation, the entropy is found to increase. The dynamical investigation simply proves that the probability that the entropy will increase is infinitely great. Purposive directive intelligence might perfectly well maintain the constancy of the entropy, or decrease it by means of compensatory actions in parts of the universe inaccessible to us. The mathematics show that in the absence of purposive direction such compensation is infinitely improbable. An example from more familiar phenomena will make this clearer. It would be theoretically possible for a whist player to deal himself a hand of thirteen trumps a thousand times in succession without any selection of the cards, but the improbability, though by no means infinite, is so great that there is, I believe, no record of its occurrence even twice in succession. A familiar example of the tendency of the entropy of a dynamical system to increase is afforded by taking as the system in question a fire and a kettle of water in contact with it. Increase of the entropy here corresponds to a transfer of heat from the fire to the water, which is not only found by observation to be the case, but is shown, by the same mathematical process as that employed in dealing with the general problem, to be infinitely probable. It is not, however, a theoretical certainty. A series of molecular changes is dynamically possible in which the entropy would decrease, that is to say, the water would give up heat to the fire, and might even freeze, while the fire became hotter. It would also be possible, but of even greater improbability, for the entropy to remain constant, in which case there would be no interchange of heat between the fire and the water, so that the water would neither be heated nor cooled. In consequence of the possibility, demonstrated by Maxwell's sorting demon illustration, of intelligent modification of molecular paths, either of these improbable

results might be effected by directive intelligence without interference with any law of nature, just as an indefinite series of all-trump hands might be dealt by conscious selection of the cards. If the placing of the kettle on a fire could be repeated an infinite number of times, the mathematics show that the occurrence of some such instances of deviation from the observed law would be infinitely probable, that is to say, practically certain. The number of times that a kettle can have been placed on a fire in the course of man's existence on the earth would, however, be insufficient to give any appreciable probability of such an event ever having occurred, much less of its having been observed. The reader with sufficient mathematical attainments will recognise that the certainty of the universe having had a beginning in time is of the same order as the certainty that the kettle of water will not retain its temperature unchanged when placed on a fire, and is of a higher order than the certainty that the water will not freeze.

The hypothesis of a universe without intelligent formative or directive power therefore leads of necessity to the conclusion that it has a beginning in time, and that it must come to an end in what we may call physical death. This state will be attained when all its organised, or available, energy has been transformed into a uniform distribution of unorganised energy, when physical change of every kind must cease. The advocates of this hypothesis are then confronted with the problem of the creation of a universe in the absence of creative intelligence.

The definite demonstration that the conclusion arrived at above is a practical necessity of the known molecular constitution of matter, and not merely an inference from the observed universality of the principle of the dissipation of energy, within the limits of our experience, requires the employment of analytical methods which will, unfortunately, limit the number of those who can hope to verify its validity for themselves. This, however, is common to many of the more recondite verities of science, and those who call them in question because they conflict with their own hypotheses must, unless they are willing to be classed with the circle squarers and perpetual motionists, first acquire the knowledge requisite for comprehending the arguments by which they are arrived at. A general idea of the nature of the argument may, however, be attained without much difficulty. Any possible series of changes of any dynamical system, such, for example, as our universe of ether and matter, may be represented mathematically by the movement of a point along a certain path, the state of the system at any instant being completely specified when we know the position and the direction of motion of the

representative point at the corresponding instant. If the point were moving in space as we know it, it would, at any point in its path, be able to move along any one of three mutually perpendicular directions, or in any intermediate direction. This would not offer a sufficient variety of paths to the moving point to enable it to represent the sequence of changes of even so simple a dynamical system as one consisting of a single particle. It is therefore necessary to suppose an ideal space to be constructed in which the point, when in any position, could travel in more than three mutually perpendicular directions or in any intermediate ones. This would increase the number of paths possible to the moving point, and by making this increase proportional to the complexity of the system, it is possible to represent all possible successions of changes of the most complicated system in this simple manner.

The reader should bear in mind that this step by step increase in the number of mutually perpendicular directions supposed to exist in the ideal space, expressing what mathematicians call the dimensions of the space, is merely a convenient method of representing the step by step increase in the complexity of the relations involved, and is found greatly to facilitate the mathematical treatment of the problem. Moreover, it appears to be the only possible way of conveying to the non-mathematical reader some faint idea of the nature of the conditions to be expressed.

In order to represent the changes in the matter in the entire universe in this manner, it would be necessary to suppose the ideal space to be of such a character that the number of mutually perpendicular directions in it would be six times as many as the number of molecules of matter in the entire universe. In Appendix K, where the proof is given in full, this number is represented by $2N'$. In order that the representative point should be capable of representing the motions of the ether in the universe as well, it would be necessary to employ an ideal space in which this number $2N'$ of mutually perpendicular directions is increased by a number $2M'$, where M' represents the number of possible distinct kinds of vibration in the ether, and is very much larger than N'. The total number of mutually perpendicular directions would then be $2N' + 2M'$, and the representative point, when in any position, would be able to move along any one of these directions, or in any intermediate direction.

Among the inconceivable infinity of paths open to the moving point, it is found that an infinitesimally small proportion of the whole would correspond to a universe which conserves for ever its state of perpetual change, while all the others would correspond

to a universe passing from a state which must have had a beginning in time to an ultimate condition of physical death, in which the energy originally existing in the matter, and necessary to its continued existence as matter, would be transferred almost entirely to the ether. In the final state the infinitesimal portion remaining in the matter, which would probably then be reduced to its constituent electrons, would be uniformly distributed throughout it, and there would also be a uniform distribution of energy throughout the ether. The universe would have passed into a state of changeless death.

At every point in the path of the representative point there would be an infinite number of wrong ways, or ways which would ultimately bring it to rest, for each one leading to an unending continuation of its motion. If the representative point moved in a wrong direction at more than an infinitesimally small proportion of the points traversed in its path, it would ultimately come to rest, when the universe would have arrived at the state of physical death, in which no further change would be possible. Let us, for a moment, imagine ourselves capable of watching the motion of such a representative point along a right path, which we must suppose that we have the power of distinguishing among the infinity of wrong ones. As we noted it from instant to instant, moving inerrantly among the infinite possibilities of error at every point, could we possibly avoid the conclusion that the path had been prescribed for it by intelligent power, and that this invisible intelligence was guiding it in its course? We must bear in mind, moreover, that the possible paths, right and wrong, as we have called them, are paths open to the representative point in accordance with the dynamical laws to which it is subject, the laws of ether and matter which govern all the changes in the material universe. The possibility of such a representation therefore proves the compatibility of such guidance with the existing laws of nature as observed. Moreover, since the combination of human intelligence which is crystallised in modern mathematical analysis is capable of constructing so simple a model to illustrate the guidance of the whole material universe by the supreme intelligence, it is no longer possible to maintain that the conception of such guidance is entirely beyond the grasp of the human mind. Such continuous external guidance is here shown to be a necessity only on the hypothesis of the endless duration of the universe. At the same time it is shown to be possible that such continuous guidance may take place without any interference with, or modification in, the observed dynamical laws of mechanics and molecular physics. Such guidance would be in no respect

analogous to the supposed interventions of the Creator suggested by Newton, and adversely criticised by Laplace, but would be strictly analogous to the guidance which, as we shall presently see, is observed to be effected, on a relatively minute scale, by all living beings.

Schemes have been propounded with a view of accounting for the established order of nature without the assumption of primal intelligence, and it would be incorrect to assert that no competent thinker has ever contemplated the possibility of developing a satisfactory one. It may, however, be asserted without fear of contradiction from anyone competent to express an opinion, that no scheme of the kind has ever been presented which would appear even superficially plausible to any but untrained minds incapable of distinguishing between argument and assertion. In the present state of scientific knowledge we are justified in maintaining that the possibility of such a scheme is unthinkable.

While comparatively few thinkers have contemplated the possibility of the elaboration of a scheme of the universe excluding the action of primal intelligence, a much larger number have maintained that the nature and qualities of such an intelligence must ever remain entirely outside the domain of human observation and inference. This *agnostic* position, as it was called, was largely prevalent during the latter portion of the past century, but far more generally among biological students than among the students of physical science. Readers to whom this would appear the only legitimate point of view would naturally regard the detailed consideration of the necessity for postulating the existence of primal intelligence as foreign to the scope of a physical treatise. They would be perfectly right, on the assumption that this intelligence is outside the region of physical investigation. It is in order to dispose of such an objection that, before proceeding with this detailed consideration, I shall now indicate the line of inquiry which will be followed out in this chapter.

I propose to introduce, in the first place, as a working hypothesis, the concept of an all-pervading *universal mind*, or omnipresent intelligence, forming an entity even more fundamental than the all-pervading ether. I shall then endeavour, by aid of the known relations existing between energy, ether, and matter, to effect a reduction to its simplest and most fundamental expression, of the observed action of mind, as exhibited by organic forms, in modifying their material environment and transmitting life to their progeny. I shall then prove that, given only the total amount of energy now present in the universe, without regard to its initial distribution, the establishment of the present order of nature can

be completely accounted for by attributing to the universal mind the possession, in more highly developed form, of powers of action similar to those exhibited in connection with existing organisms, together with the intelligence necessary for their exercise. Finally, I shall argue that the possession of such intelligence by the universal mind is a legitimate inference from the highest manifestations of human intelligence considered as embryonic forms originating from the universal mind.

The hypothesis affords a satisfactory explanation of the otherwise inexplicable power exhibited by the human mind of comprehendingly reading in the book of nature, which has always appeared to me as the most wonderful of all the wondrous aspects of that awe-inspiring volume. This was very beautifully expressed by Professor W. K. Clifford in the second number of *The Nineteenth Century*, where, referring to Dr. Martineau's theistic argument based upon the existence of moral law and the evidence of design in nature, he wrote as follows:—

> "I fully admit that the theistic hypothesis, so grounded, and considered apart from objections elsewhere arising, is a reasonable hypothesis and an explanation of the facts. The idea of an external conscious being is unavoidably suggested, as it seems to me, by the categorical imperative of the moral sense; and, moreover, in a way quite independent, by the aspect of nature, which seems to answer our questionings with an intelligence akin to our own."

The unity and uniformity of nature, which finds its expression in the principle of continuity, could have no conceivable meaning in the absence of primal intelligence, and it would be difficult to formulate a better expression of this principle than the one given by Balfour Stewart and Tait:—[1]

> "Assuming the existence of a supreme Governor of the universe, the principle of continuity may be said to be the definite expression in words of our trust that He will not put us to permanent intellectual confusion, and we can easily conceive similar expressions of trust with reference to the other faculties of man."

How else, again, than by the operation of primal intelligence can the origin of the electrons become conceivable? According to the conceptions of the electromagnetic scheme which I have endeavoured to develop in this volume, matter, in the numerous forms now known to us, may be ultimately analysed into two classes of strain forms in the ether. Every individual form in either of these classes resembles every other form in its class with such exactness that all the refinements of modern instruments of

[1] *Unseen Universe*, 6th edition, p. 88.

precision, aided by the even more wonderful refinements of mathematical analysis, have failed to reveal the smallest indication of variability. The two classes of electrons, in fact, bear the obvious impress of manufactured articles of a perfection of which the most highly trained mind can form but an inadequate conception. There is no known physical process, undirected by intelligence, by which the formation of even one of these electrons could be conceived; nor, with the exception of the infinitely improbable coalescence of two electrons of different classes, is there any such process by which they can conceivably be destroyed, provided the fluidity of the ether is perfect. If, on the other hand, the ether possesses the slightest infinitesimal trace of viscosity, the whole material universe must ultimately dissolve into formless ether without leaving a vestige behind, just as a smoke ring dissolves in the air. The analysis of material molecules into electrons was unthought of in the year 1873, when Clerk-Maxwell delivered his famous lecture on Molecules before the Bradford meeting of the British Association. We should, however, be perfectly justified in applying to the electrons the concluding words of that memorable address:—

"They continue this day as they were created, perfect in number and measure and weight, and from the ineffaceable characters impressed on them we may learn that those aspirations after accuracy in measurement, truth in statement, and justice in action, which we reckon among our noblest attributes as men, are ours because they are essential constituents of the image of Him who in the beginning created, not only the heaven and the earth, but the materials of which heaven and earth consist."

The hypothesis of the universal mind leads to the concept of a third model of the universe consisting of mind, energy, ether, and matter, determining, but not determined by, the molecular scheme. Just as the molecular scheme must not be regarded as a sink for the disposal of mechanical difficulties, so the mental scheme, as it may be called, must not be regarded as a sink for the disposal of molecular difficulties.

The relations between the mechanical, molecular, and mental models have been very lucidly stated by Sir Joseph Larmor,[1] who observes, in reference to the phenomena of magnetism,

"it would appear that there can be an unlimited amount of molecular structure and function in a given system which is unconnected with any mechanical effect occurring in that system treated as continuous matter. This is because, whether we view it as an independent principle or as a corollary from the principle of action, these mechanical

[1] *Æther and Matter*, p. 287.

relations are, from their very nature, determined by analytical functions of configuration and their first and second gradients alone [that is to say, by their first and second rates of change relative either to time or space]: if higher gradients also came in, the statement would no longer hold. The processes by which our conception of the uniformity of nature is obtained essentially involve averaging of effects, and lose their efficacy long before the individual molecule is reached. Mechanical determinateness thus need not involve molecular determinateness: then why should either of them involve determination in the entirely distinct province of vital activity? Moreover, mechanical science has to do with systems in being: it does not avail to trace the circumstances of growth or structural change even in inorganic material. What happens when two gaseous molecules unite to form a compound molecule is unknown except from the slight indirect indications of spectrum analysis. Now, all initiation of organic activity seems to involve structural change, not merely mechanical disturbance, and is, in so far, outside the domain of mechanical laws. But the activities of an organism treated as a permanent system—such, for example, as propagation of nervous impulse—are likely enough, when once they are started, to be of the nature of the interactions of matter in bulk, so that it is legitimate to seek for them a mechanical correlation. Every vital process may conceivably thus be correlated with a mechanical process, as to its progress, just to that extent to which it is possible experimentally to follow it, without lending any countenance to a theory that would place its initiation under the control of any such system of mechanical relations. In other terms, there is room for complete mechanical co-ordination of all the functions of an organism, treated as an existing material system, without requiring any admission that similar principles are supreme in the more remote and infinitely complex phenomena concerned in growth and decay of structure."

The contention that the human mind is necessarily confined to a knowledge of phenomena, that is to say, material phenomena, only, is very ably dealt with by Professor Max Müller in the following paragraphs:—[1]

"I need hardly say that the very name phenomenal or apparent implies that there is something that appears, something of which we can therefore predicate that it appears, something that seems to be, that is relative to us, and so far, but so far only, known to us. That which appears is, before it appears, unknown to us, but it becomes known to us, in the only way in which it can be known, that is, by its appearance, by its phenomenal manifestations, by its becoming an object of human knowledge. It is known to us as that without which the phenomenal would be impossible, nay, unthinkable. That without which the phenomenal would be unthinkable is sometimes called the noumenal, the real, the absolute, and if we call its absence unthinkable, we imply that there are certain forms of our thought from which our

[1] "Why I am not an Agnostic," *Last Essays*, pp. 349–356.

phenomenal knowledge cannot escape, the well-known Kantian forms of intuition and understanding. These, as Kant has shown, cannot be the mere result of phenomenal experience, because they possess a character of necessity which no phenomenal experience can ever claim. . . ."[1]

This is exemplified by pointing out that

"if we think at all, we must submit to the law of causality, a category of our understanding without which even the simplest phenomenal knowledge would be impossible. We never see a horse, we are only aware of certain states of our own consciousness, produced through our senses; but that these affections presuppose a cause, or, as we call it, an object outside us, is due to that law of casuality within us which we must obey, whether we like it or not. If then we have to recognise in every single object of our phenomenal knowledge a something or a power which manifests itself in it, and which we know, and can only know, through its phenomenal manifestation, we have also to acknowledge a power that manifests itself in the whole universe. We may call that power unknown or inscrutable, but we may also call it the best known, because all our knowledge is derived from a scrutiny of its phenomenal manifestations. That it is, we know; what it is by itself, that is, out of relation to us or unknown by us, of course we cannot know, as little as we can eat our cake and have it; but we do know that without it the manifest or phenomenal universe would be impossible.

Again:—

"I cannot help discovering in the universe an all-pervading causality or a reason for everything; for, even when in my phenomenal ignorance I do not yet know a reason for this or that, I am forced to admit that there exists some such reason; I feel bound to admit it, because to a mind like ours nothing can exist without a sufficient reason. But how do I know that? Here is the point where I cease to be an Agnostic. I do not know it from experience, and yet I know it with a certainty greater than any which experience could give." . . . "We ourselves, after thousands of years of physical and metaphysical research, can say no more than that there is Νοῦς, that there is mind and reason in nature. *Sa Majesté le Hasard* has long been dethroned in scientific studies, and neither natural selection, nor struggle for life, nor the influence of environment, or any other *aliases* of it, will account for the *Logos*, the thought, which with its thousand eyes looks at us through the trans-

[1] I quite agree with Bertrand Russell (*Principles of Mathematics*, vol. i. p. 4) that Kant's contention that mathematical reasoning is not strictly formal, but always uses intuitions, *i.e.* the *a priori* knowledge of space and time, is no longer defensible in view of the progress of symbolic logic, one of the results of which has been to show that the whole of pure mathematics may be established upon an arithmetical basis, and that our conceptions of time, and of the space in which we exist, are not intuitions, but physical concepts derived from experience. This does not, however, invalidate the above argument, or the example in the following paragraph, which is one of two given by Max Müller.

parent curtain of nature and calls for thoughtful recognition from the *Logos* within us. If any philosopher can persuade himself that the true and well-ordered genera of nature are the result of mechanical causes, whatever name he may give them, he moves in a world altogether different from my own. He belongs to a period of thought antecedent to Anaxagoras. To Plato these genera were ideas ; to the Peripatetics they were words or *logoi* ; to both they were manifestations of thought. Unless these thoughts had existed previous to their manifestation or individualisation in the phenomenal world, the human mind could never have discovered them there and named them. We ought not to say any longer, in the language of the childhood of the race, 'In the beginning God created heaven and earth.' As Christians we have to say, in the language of St John and his Platonic and Gnostic predecessors, 'In the beginning there was the *Logos*.' If we call that *Logos* the Son, and if we speak of a Father whom no one knows but the Son, the so-called *Deus ante intellectum*, we are using human language, but if we know that all human language is metaphorical, we shall never attempt to force these words into a narrow literal meaning. To do so is to create mythology, and with it all its concomitant dangers. What lies beyond the curtain of these words is, in fact, the legitimate realm of *Agnoia* or Agnosticism. But all that lies on this side of the curtain is our domain, the domain of language and afterwards of science, which in the chaos of phenomena has discovered, and with every new generation of Aristotles, Bacons, and Darwins is bent on discovering more and more, a hidden *Cosmos*, or the reflex of that *Logos*, without which nature would be illogical, irrational, chaotic, and existing by accident only, not by the will of a rational Power. Call that Power the Father, or call it a Person, and you neither gain nor lose anything, for these words also are metaphorical only, and what constitutes the personal element in man or any other living being is as unknown to us as what constitutes the personal element in the author, the thinker, the speaker, or creator of the *logoi*. All I maintain is, that if we ever speak of a *Logos* and of *logoi*, and understand clearly what we mean by these words, we can no longer say that in the beginning there was protoplasm, and that the whole world was evolved from it by purely mechanical or external agencies. If we have once recognised in all the genera or generations of the material world, not simply the unknown, or a substance and power that is inscrutable, but the thoughts and will of a mind, that mind, so far from being inscrutable, undergoes a constant scrutiny in its endless manifestations at the hand of human science. It is, in fact, the one subject of all our knowledge, from the first attempts at roughly grouping and naming it to the latest efforts of scientific research, intended to classify, to comprehend, and understand it. The whole of our knowledge of nature becomes thus a recognition of the *logoi* of nature by the *Logos* of ourselves. Each genus becomes a *logos*, an eternal thought or an eternal word ; nay, it seems to follow from this that there is in nature no room for anything but genera, no room for species or εἴδη in the proper sense of these terms. Here we see how the Science of Language becomes the Science of Thought. If it is unity of origin which constitutes a genus, true science knows indeed of individuals which represent a genus, but not of species, though for practical purposes the human mind may give that name to varieties in their more or less inheritable and

permanent form; such varieties being in reality no more than the necessary consequence of individualisation and manifoldness." . . . "I am not aware that on my way to this Gnosis I have availed myself of anything but the facts of our direct consciousness, and the conclusions that can be logically deduced therefrom. Without these two authorities I do not feel bound to accept any testimony, whether revealed or unrevealed." . . . "In one sense I hope I am, and always have been, an Agnostic, that is, in relying on nothing but historical facts and in following reason as far as it will take us in matters of the intellect, and in never pretending that conclusions are certain which are not demonstrated or demonstrable. This attitude of the mind has always been recognised as the *conditio sine qua non* of all philosophy. If in future it is to be called Agnosticism, then I am a true Agnostic; but if Agnosticism excludes a recognition of an eternal reason pervading the natural and the moral world, if to postulate a rational cause for a rational universe is called Gnosticism, then I am a Gnostic, and a humble follower of the greatest thinkers of our race, from Plato and the author of the Fourth Gospel to Kant and Hegel."

The conclusion that mind is the ultimate determinative power in nature, although derived from other than the purely physical considerations with which alone we are concerned, is expressed by Cardinal Newman[1] in the following beautiful and forcible terms:—

"The agency then which has kept up and keeps up the general laws of nature, energising at once in Sirius and on the earth, and on the earth in its primary period as well as in the nineteenth century, must be Mind and nothing else, and Mind at least as wide and as enduring in its living action as the immeasurable ages and spaces of the universe on which that agency has left its traces."

That the necessity of a primal intelligence strongly impressed itself upon the mind of the greatest of biologists, Charles Darwin, is shown in his letter to Asa Gray, written on 22nd May 1860, in which he observed:—

"I cannot anyhow be contented to view this wonderful universe, and especially the nature of man, and to conclude that everything is the result of brute force. I am inclined to look at everything as resulting from designed laws, with the details, whether good or bad, left to the working out of what we call chance. Not that this notion at all satisfies me."

In a later letter (26th November 1860) to the same correspondent he wrote:—

"I am conscious that I am in an utterly hopeless muddle. I cannot think that the world, as we see it, is the result of chance; and yet I cannot look at each separate thing as the result of design."

---

[1] *Grammar of Assent*, p. 70.

From our present point of view, a simple and natural solution of this difficulty suggests itself. We are compelled, by our natural limitations, to approach the consideration of the molecular scheme from the standpoint of the mechanical scheme, and many of the problems of the former can only be attacked by the statistical methods appertaining to the calculus of probabilities. The more minute details of such problems are therefore beyond the reach of our present methods, except in so far as the laws of chance are applicable. That is the only sense in which the physicist can admit that these details are left to the working out of chance. The mental scheme, again, can only be approached through the molecular scheme, and must therefore exhibit to us all the apparent indeterminateness of the latter, together with that which is inherent in its relation to the molecular scheme itself, and would therefore remain, even should our methods of investigation attain a perfection sufficient completely to unfold the molecular scheme to our apprehension. This in nowise prevents the complete determination of the mechanical scheme by the molecular one, or of the latter by the mental scheme.

There is no way of evading the conclusion that a determining cause must be sought for beyond the molecular scheme. There is one and only one such cause known to us—our own will or mind; and the fundamental principles of scientific investigation lead us, therefore, to seek in the extension of mind for the determination of the molecular scheme, and further, of the whole order of nature. We find that the mental scheme, introduced simply as a working hypothesis, proves satisfactory at every point where the molecular scheme is found to be insufficient, and the attempt to ignore it in the development of any scheme attempting to account for the order of nature, will invariably be found to necessitate its introduction in some disguised and unscientific manner, which very commonly takes the form of personifying natural law, one of the worst of pseudo-scientific absurdities.

"What," Huxley asks,[1] "is the dire necessity and 'iron law' under which men groan? Truly, most gratuitously invented bugbears. I suppose, if there be an 'iron law,' it is that of gravitation; and if there be a physical necessity, it is that a stone, unsupported, must fall to the ground. But what is all we really know, and can know, about the latter phenomena? Simply that, in all human experience, stones have fallen to the ground under these conditions; that we have not the smallest reason for believing that any stone so circumstanced will not fall to the ground; and that we have, on the contrary, every reason to believe that it will so fall. It is very convenient to indicate that all

[1] "Physical Basis of Life," *Lay Sermons*, pp. 124, 125.

the conditions of belief have been fulfilled in this case, by calling the statement that unsupported stones will fall to the ground 'a law of nature.' But when, as commonly happens, we change *will* into *must*, we introduce an idea of necessity which most assuredly does not lie in the observed facts, and has no warranty that I can discover elsewhere. For my part, I utterly repudiate and anathematise the intruder. Fact I know, and Law I know; but what is this Necessity, save an empty shadow of my own mind's throwing? But if it is certain that we can have no knowledge of the nature of either matter or spirit, and that the notion of necessity is something illegitimately thrust into the perfectly legitimate conception of law, the materialistic position that there is nothing in the world but matter, force, and necessity is as utterly devoid of justification as the most baseless of theological dogmas."

The pure idealist accepts mind as the fundamental entity through which alone the representation which we call the external world is made possible, and then takes the further and entirely unwarranted step of maintaining that matter can have no real existence, and that the external world has no reality other than its representation in his consciousness. The pure materialist, on the other hand, maintains that matter and motion are the only realities, and that mind has no independent existence of its own—a still more unphilosophical conclusion. The purely idealist position might serve as a working hypothesis for a philosopher who had never within his memory been in communication with other minds, for some of the strongest arguments against the existence of matter are equally cogent against the existence of other minds. The materialist position appears to me to be vitiated at its origin by its refusal to admit that the only immediate and absolutely certain knowledge possible to us is that of our own states of consciousness. Our knowledge of matter, and of the external world generally, including our own bodies, is, I consider, necessarily mediate knowledge consisting entirely of our interpretations of these states of consciousness. This is very clearly explained by Huxley in his discourse on Descartes.[1]

Referring to Descartes, he observes :—

"As the record of his progress tells us, he was obliged to confess that life is full of delusions: that authority may err; that testimony may be false or mistaken; that reason lands us in endless fallacies; that memory is often as little trustworthy as hope; that the evidence of the very senses may be misunderstood; that dreams are real as long as they last, and that what we call reality may be a long and restless dream. Nay, it is conceivable that some powerful and malicious being may find his pleasure in deluding us, and in making us believe the thing which

[1] *Lay Sermons*, p. 283.

is not, every moment of our lives. What, then, is certain? What even, if such a being exists, is beyond the reach of his powers of delusion? Why, the fact that the thought, the present consciousness, exists. Our thoughts may be delusive, but they cannot be fictitious. As thoughts, they are real and existent, and the cleverest deceiver cannot make them otherwise.

"Thus, thought is existence. More than that, so far as we are concerned, existence is thought, all our conceptions of existence being some kind or other of thought. Do not for a moment suppose that these are mere paradoxes or subtleties. A little reflection upon the commonest facts proves them to be irrefragable truths. For example, I take up a marble, and I find it to be a red, round, hard single body. We call the redness, the roundness, the hardness, and the singleness 'qualities' of the marble, and it sounds, at first, the height of absurdity to say that all these qualities are modes of our own consciousness, which cannot even be conceived to exist in the marble. But consider the redness, to begin with. How does the sensation of redness arise? The waves of a certain very attenuated matter [this was the view then taken of the ether] the particles of which are vibrating with vast rapidity, but with very different velocities, strike upon the marble, and those which vibrate with one particular velocity are thrown off from its surface in all directions. The optical apparatus of the eye gathers some of these together, and gives them such a course that they impinge upon the surface of the retina, which is a singularly delicate apparatus, connected with the termination of the fibres of the optic nerve. The impulses of the attenuated matter, or ether, affect this apparatus and the fibres of the optic nerve in a certain way, and the change in the fibres of the optic nerve produces yet other changes in the brain; and these, in some fashion unknown to us, give rise to the feeling or consciousness of redness. If the marble could remain unchanged, and either the rate of vibration of the ether or the nature of the retina could be altered, the marble would seem not red, but some other colour. There are many people who are what is called colour-blind, being unable to distinguish one colour from another. Such an one might declare our marble to be green; and he would be quite as right in saying that it is green as we are in declaring it to be red. But then, as the marble cannot, in itself, be both green and red at the same time, this shows that the quality 'redness' must be in our consciousness and not in the marble.

"In like manner it is easy to see that the roundness and the hardness are forms of our consciousness, belonging to the groups which we call sensations of sight and touch. If the surface of the cornea were cylindrical, we should have a very different notion of a round body from that which we possess now; and if the strength of the fabric, and the force of the muscles of the body, were increased a hundredfold, our marble would seem to be as soft as a pellet of bread-crumbs.

"Not only is it obvious that all these qualities are in us, but, if you will make the attempt, you will find it quite impossible to conceive of 'blueness,' 'roundness,' and 'hardness' as existing without reference to some such consciousness as our own. It may seem strange to say that even the 'singleness' of the marble is relative to us; but extremely simple experiments will show that such is veritably the case, and that our two most trustworthy senses may be made to contradict

one another on this very point. Hold the marble between the finger and thumb and look at it in the ordinary way. Sight and touch agree that it is single. Now squint, and sight tells you that there are two marbles, while touch asserts that there is only one. Next, return the eyes to their natural position, and having the forefinger and the middle finger crossed, put the marble between their tips. Then touch will declare that there are two marbles, while sight says that there is only one; and touch claims our belief, when we attend to it, just as imperatively as sight does.

"But it may be said the marble takes up a certain space which could not be occupied at the same time by anything else. In other words, the marble has the primary quality of matter—extension. Surely this quality must be in the thing, and not in our minds? But the reply must still be: whatever may or may not exist in the thing, all that we know of these qualities is a state of consciousness. What we call extension is a consciousness of a relation between two or more affections of the sense of sight or of touch. And it is wholly inconceivable that what we call extension should exist independently of such consciousness as our own. Whether, notwithstanding this inconceivability, it does so exist or not is a point on which I offer no opinion.

"Thus, whatever our marble may be in itself, all that we can know of it is in the shape of a bundle of our own consciousnesses.

"Nor is our knowledge of everything we know or feel more or less than a knowledge of states of consciousness. And our whole life is made up of such states. Some of these states we refer to a cause we call 'self'; others to a cause or causes which may be comprehended under the title of 'not-self.' But neither of the existence of 'self' nor of that of 'not-self,' have we, nor can we by any possibility have, any such unquestionable and immediate certainty as we have of the states of consciousness which we consider to be their effects. They are not immediately observed facts, but results of the application of the law of causation to those facts. Strictly speaking, the existence of a 'self' and of a 'not-self' are hypotheses by which we account for the facts of consciousness. They stand upon the same footing as the belief in the general trustworthiness of memory, and in the general constancy of the order of nature—as hypothetical assumptions which cannot be proved or known with that highest degree of certainty which is given by immediate consciousness; but which, nevertheless, are of the highest practical value, inasmuch as the conclusions logically drawn from them are always verified by experience."

Again, p. 286:—

"Descartes, having commenced by declaring doubt to be a duty, found certainty in consciousness alone; and the necessary outcome of his views is what may properly be termed Idealism; namely, that whatever the universe may be, all that we can know of it is the picture presented by consciousness. This picture may be a true likeness—though how this can be is inconceivable; or it may have no more resemblance to its cause than one of Bach's fugues has to the person who is playing it; or than a piece of poetry has to the mouth and lips of a reciter. It is enough for all the practical purposes of human

existence if we find that our trust in the representations of consciousness is verified by results; and that, by their help, we are enabled 'to walk surefootedly in this life.'

"Thus the method, or path which leads to truth, indicated by Descartes, takes us straight to the Critical Idealism of his great successor Kant. It is that Idealism which declares the ultimate fact of all knowledge to be a consciousness, or, in other words, a mental phenomenon; and therefore affirms the highest of all certainties, and indeed the only absolute certainty, to be the existence of mind. But it is also that Idealism which refuses to make any assertions, either positive or negative, as to what lies beyond consciousness. It accuses the subtle Berkeley of stepping beyond the limits of knowledge when he declared that a substance of matter does not exist; and of illogicality for not seeing that the arguments which he supposed demolished the existence of matter were equally destructive to the existence of soul."

Again, p. 287:—

"The path indicated and followed by Descartes which we have hitherto been treading, leads through doubt to that Critical Idealism which lies at the heart of modern metaphysical thought. But the *Discourse* shows us another, and apparently very different, path, which leads, quite as definitely, to that correlation of all the phenomena of the universe with matter and motion which lies at the heart of modern physical thought, and which most people call Materialism."

After pointing out the good work accomplished by the Materialists in applying the conceptions of mechanics and molecular physics to the explanation of the physiology of living bodies, Huxley observes, p. 296:—

"When the Materialists stray beyond the borders of their path and begin to talk about there being nothing else in the universe but Matter and Force and Necessary Laws, and all the rest of their 'grenadiers,' I decline to follow them." . . . "'Matter' and 'Force' are, as far as we know, mere names for certain forms of consciousness. 'Necessary' means that of which we cannot conceive the contrary. 'Law' means a rule which we have always found to hold good, and which we expect always will hold good. Thus it is an indisputable truth that what we call the material world is only known to us under the forms of the ideal world; and, as Descartes tells us, our knowledge of the soul is more intimate and more certain than our knowledge of the body. If I say that impenetrability is a property of matter, all that I can really mean is that the consciousness I call extension, and the consciousness I call resistance, constantly accompany one another. Why and how they are thus related is a mystery. And if I say that thought is a property of matter, all that I can mean is that, actually or possibly, the consciousness of extension and that of resistance accompany all other sorts of consciousness. But, as in the former case, why they are thus associated is an insoluble mystery.

"From all this it follows that what I may term legitimate materialism that is, the extension of the conceptions and of the methods of physical

science to the highest as well as the lowest phenomena of vitality, is neither more nor less than a sort of shorthand Idealism; and that Descartes' two paths meet at the summit of the mountain, though they set out on opposite sides of it.

"The reconciliation of physics and metaphysics lies in the acknowledgment of faults on both sides; in the confession by physics that all the phenomena of nature are, in their ultimate analysis, known to us only as facts of consciousness; in the admission by metaphysics that the facts of consciousness are, practically, interpretable only by the methods and the formulæ of physics; and finally, in the observance by both metaphysical and physical thinkers of Descartes' maxim—assent to no proposition the matter of which is not so clear and distinct that it cannot be doubted."

At the first dawn of consciousness in a child, whether it occurs before birth, or afterwards, say when it first opens its eyes upon the world, it can know nothing but its state of consciousness at the moment. Very soon it will begin to have, in addition, the memory of states of consciousness through which it has passed, it will become conscious of desire and will, and the memories of previous states of its consciousness will gradually be built up into a more or less ordered system, and it will begin to acquire a knowledge of the external world long before it is capable of making a distinction between the self and the not-self. These early experiences are completely forgotten long before the age of reason is attained, and with it the power of making a distinction between these two entities. A very much higher intellectual stage must be reached before we can begin to question our consciousness as to the grounds on which this distinction is made. When we do this, we see that it is impossible to question the existence of the self. Theoretically we may question the existence of the not-self or external world; that is to say, each one of us may ask himself the question: Am I the all, or is there something outside of myself? The question is, however, answered as soon as it is framed, for the limitations of the power of the self to order its own states of consciousness compel the belief in something outside of, and conditioning, the self. We may maintain, with the pure Idealist, that there is no material external world, but we are then compelled to admit the existence of an external power which orders our states of consciousness as if there were an external world. That is to say, in attempting to avoid the hypothesis directly presented by consciousness as the interpretation of its states, we reintroduce it in another form. It must always be borne in mind that scientific knowledge consists in the establishment of relations between the phenomena presented by our consciousness. The whole trend of progress in such knowledge is towards the complete unification of all phenomena as manifesta-

tions of one underlying ultimate reality which philosophers call the Absolute. All that we can ever conceivably learn with regard to this is the nature of its manifestations as presented to our consciousness. As a phenomenal reality, then, the existence of a world external to ourselves cannot be questioned. Now this external world presents itself in a certain ordered relation with our states of consciousness. We usually interpret this by saying that it affects our states of consciousness, and that these, when they take the form of will, have, in their turn, the power of affecting the external world.

The science of physics, including physiology, the physics of living matter, leads us to the conclusion that all changes in consciousness are accompanied by changes in the grey matter of the brain; and that certain changes in the grey matter of certain portions of the brain are, conversely, under normal conditions, invariably accompanied by changes of consciousness. In all observed cases of the concomitance of conscious and material phenomena, physical science has enabled us to explain with more or less completeness all the stages of the relation except that between the brain changes and the changes in consciousness. If this stage could be explained, we should arrive at a solution of the problem of the relation between mind and matter. I shall use the term mind to denote the entity which forms the basis of all the phenomena of consciousness in man and the higher animals, and of all those purposive actions which enable the lower as well as the higher organic forms to modify the relations between themselves and their material environment. This modifying power is the criterion of life, which is defined by Herbert Spencer as a capacity for "the continuous adjustment of internal relations to external relations."[1] Three distinct hypotheses have been propounded by psychologists for the representation of the observed correspondences of mental and material phenomena. The first assumes a mutual causal relation between the two classes, so that phenomena of either may be considered as directly giving rise to phenomena of the other. The second admits action by the external world upon the consciousness, but denies any reciprocal action of consciousness upon the external world. The third maintains that the two sets of phenomena are merely concomitant, without any causal relation at all. The first hypothesis is the one directly suggested by the facts of consciousness, and is the one which regulates the actions of every rational human being, even if he gives a theoretical assent to either of the others. The second and third represent the

[1] *Principles of Biology*, vol. i. p. 74.

consciousness of a man or animal as resembling a mirror in which the external world is pictured, but which has no power of influencing the nature or succession of the pictures. The difference between the second and third is that, according to the former, the pictures are produced by causal action of the material environment, while, according to the latter, they arise from other, and unexplained, sources. Both these hypotheses make the consciousness of every man and animal incapable of exercising any activity in the universe, and are, for this reason, opposed to the trend of physical science towards the correlation of all phenomena into one all-embracing unity. Moreover, while affording a representation of some of the facts of consciousness, they entirely fail to account for the sense of possessing a power of voluntary action, of selection between various alternatives, which presents itself in the consciousness of each one of us. If we were limited to these three alternatives we should, therefore, be led to prefer the first as the only one which is in complete accordance with the observed facts. It has, moreover, the additional advantage of being in no way incompatible with the unity of nature indicated by the mechanical and molecular schemes.

A more fundamental, and far more satisfactory, method of attack is to follow the same course which has led to the solution of so many apparently insoluble problems in the mechanical and molecular schemes. Whenever we have attempted to correlate two series of physical phenomena, even when the relation between them is both intimate and obvious, we have found that only a very imperfect correlation could be effected by isolating the two series, and regarding them as connected by a direct causal relation independent of the remainder of the complex of which they form a part. On the other hand, we have ever found it entirely beyond our powers to attack the problem immediately in all its generality. The direct contemplation of the All simply overwhelms us with its vastness. We have found that the only road to success lies in the construction of successive working models, of gradually increasing complexity, as larger portions of the All are brought within their purview, but the complete correlation of even the simplest series can only be dimly perceived by the eye of scientific faith, as the outlines of a cloud-shrouded peak to which we are ever approaching, though its summit must ever remain above our reach. When the problem of the relation between mind and matter is approached in this manner, we recognise the first hypothesis, postulating a mutual causal relation between them, as the necessary step towards arriving at a first approximation. The final complete correlation, as in the case of the phenomena of the mechanical and molecular schemes, would require the powers of the universal

mind for its complete grasp, and all our progress in physical generalisation may be regarded as attempts—and feeble attempts they must ever be—to approach towards the all-embracing and all-comprehending point of view of the universal mind.

This method of regarding the problem appears to me to throw an entirely new light on that fundamental philosophical difficulty, the question of the freedom or determinism of the will. In a mental scheme determined by an all-wise and all-prescient universal mind, there is no place for the free, unfettered action of independent units of mind. On the other hand, there is no more certainly ascertained fact of consciousness than the conviction that there is a certain freedom of choice open to each individual unit of will. We have seen that there is every reason to believe that the mechanical scheme is completely determined by the molecular scheme; that is to say, that, to an intelligence capable of grasping the latter in its totality, the former would be uniquely determined. We have also seen that there is no determinateness in the other direction, in that any one molecular representation of a mechanical phenomenon, shown to be possible, involves the possibility of an infinite number of others. Our present physical knowledge, therefore, not merely admits of, but suggests, the co-existence of complete freedom of the will, relatively to the molecular scheme, with its complete determinism in terms of the mental scheme. Another consideration which may assist in the apprehension of this point of view is that every individual unit of mind must be regarded as a mere differentiation or outcome of the universal mind, and as connected, through the latter, with every other unit, just as every individual material atom is a mere differentiation of a portion of the ether, and is connected, through the latter, with every other atom. It does not, therefore, appear antecedently improbable that the more highly developed among these units should be conscious of sharing in the activity of the greater unity of which they form a part, and at the same time, not fully conscious of the restraint necessarily involved in such incorporation.

We now see that a first approximation to the solution of the complete interaction between mind and its material environment, if attainable at all, is to be sought in the interaction between an individual human mind and its immediate material environment, the grey matter of the individual brain, both mind and brain being considered as independent units, that is to say, without regard to their respective relations with the universal mind and the material environment. This is the starting point, according to the working hypothesis of the universal mind, for the inquiries of both physiologists and psychologists. The former have fully

established the correspondence, already referred to, between changes in consciousness and physical changes in the grey matter of the brain, and we have seen that we are entitled, as a first approximation, to consider this correspondence as arising from interaction between the individual mind and the individual brain; that is to say, the correspondence may be considered as establishing the existence of interaction between mind and matter. The former is an entity of which we are unable even to attempt any further analysis with a view to its simplification, but this is not the case with matter. According to the view advanced in the present volume, the atoms of matter consist of aggregates of electrons which are themselves simply strain forms in the ether. Material atoms are therefore simply differentiations of portions of the original formless ether arising from certain distributions of motion, to speak in terms of material analogy, or, in more general or abstract terms, certain energy configurations. The ultimate basis of matter, as far as it is in any respect accessible to observation, is the formless ether, in which as yet no strain forms exist, but which is imbued with energy of spin in every smallest portion into which we could, by the utmost stretch of our imagination, conceive it to be divided. That is to say, ether which, though as yet formless, is prepared for the formation of strain forms, as soon as the requisite formative power shall come into action. In the formless state the energy of spin forms a uniform volume, or spatial, distribution.

Let us now take, in imagination, the one further step conceivable, and suppose this uniform volume distribution of energy to be abstracted. Nothing can then remain but an absolutely quiescent ether containing no energy. It is impossible for us to form any conception of such a substance, for it would have no properties by which its presence could become known to us, these properties arising entirely from the energy distribution. It therefore has no phenomenal existence; whether or not it exists in the metaphysical sense, as a vehicle for carrying energy, we have no means of determining, for it can have no part in any known material phenomena, all of which are due to changes in energy distribution. All the phenomena of the material universe may therefore be considered as arising solely from changes in energy distribution. That is to say, energy is the sole ultimate phenomenal basis of matter.

We saw in Chapter XXII. that from the standpoint of the molecular scheme the principle of the conservation of energy was deposed from the commanding position of a fundamental principle which it appeared to occupy in the mechanical scheme. From the present still more fundamental point of view, it assumes an

even more commanding position than in the mechanical scheme, since the total energy of a dynamical system is now seen to form its sole essential ultimate content, which can only be altered by adding to, or subtracting from, that content. The principle of its conservation, therefore, presents itself as the ultimate fundamental law of all material phenomena.

Since, from the phenomenal point of view at any rate, energy is the sole ultimate basis of matter, it follows that all interactions between mind and matter are capable of finding their ultimate expression in the form of interactions between mind and energy. The observed correlation of mental and material phenomena, therefore, definitely demonstrates the power of the human mind, and of the minds of other living beings, to influence, and be influenced by, changes in the distribution of energy in their material environment. Now, given the total energy of any material system, even if the system be the whole material universe, the building up of that system from any initial distribution of its energy, and its transformation into any other possible system with the same total energy content, requires only this power of influencing the distribution of its energy, combined with the intelligence necessary for its appropriate purposive direction. The formation of the existing material universe, together with the establishment and maintenance of all known physical laws, can therefore be accounted for, on the hypothesis of the universal mind, by attributing to that mind the possession of powers differing only in degree, and not in kind, from the observed powers of living beings, together with the intelligence adequate for their purposive use. If we attempt to compare the rudimentary intelligence exhibited by the earliest and lowest organic forms with the most highly developed human minds, regarding the former as embryonic forms of the latter, and then picture to ourselves the combined intelligence of the human race and all other living forms as but an embryonic representation of the intelligence of the universal mind, we may form something at least approaching a concept of the intelligence to be attributed to the latter. In so doing we have, moreover, in no way overstepped the legitimate physical paths of inference from observed facts.

The conclusion that an intelligently directed power of influencing the distribution of energy is, from the present scientific view of the nature of matter, sufficient to account for all material phenomena, and that the attribution of this power to the primal intelligence is a legitimate inference from observation, both confirms and extends, and is, moreover, confirmed by, Maxwell's demonstration (see p. 234) of the possibility of such directive action in a special case. Whether any organic forms are capable of

effecting the transformation, considered by Maxwell, of unavailable into available energy, is at present an undetermined question. The second of the three following conclusions relating to the dissipation of energy formulated by Lord Kelvin, in a short note in 1852,[1] expressed the opinion that they are not:—

"(1) There is at present in the material world a universal tendency to the dissipation of mechanical energy.

"(2) Any *restoration* of mechanical energy, without more than an equivalent dissipation, is impossible by inanimate material processes, and is probably never effected by means of organised matter, either endowed with vegetable life or subject to the will of an animated creature.

"(3) Within a finite period of time past, the earth must have been, and within a finite period of time to come, the earth must again be, unfit for the habitation of man as at present constituted, unless operations have been, or are to be, performed which are impossible under the laws to which the known operations going on at present in the material world are subject."

According to Sir Joseph Larmor,[2] Lord Kelvin still retained, in 1892, the decided opinion expressed in (2) on the question

"whether the refinements of minute structure and adaptation in vital organisms may permit departure from the law of dissipation, which is known to be inflexible in the inorganic world, by utilising to some extent diffuse thermal energy for the production of vital mechanical power."

Helmholtz preferred to the end of his life to leave this question open, as is shown by a passage in his *Lectures on Heat*, which were published after his death. The question is one of great interest, but the nature of the answer is not of any fundamental import to the present argument.

It might appear a natural extension of the conclusions to which we have been led, to attribute to the universal mind the power of creating, that is to say, calling into existence, the energy of the universe, in addition to the power of changing its distribution. I am not, however, aware of any physical basis for making such an extension. Such a basis would be provided if it could be shown that any organism possessed the power of creating energy, but all the evidence at present available tends to the conclusion that living organisms can neither create nor destroy energy, although they are able to effect changes in its distribution. It is perfectly reasonable to suppose that the universal mind may possess further powers beyond those exhibited by minds

[1] *Mathematical and Physical Papers*, vol. i. p. 514.

[2] Obituary Notice of Lord Kelvin, *Roy. Soc. Proc.*, A, vol. lxxxi., p. xxxvi., 1908.

accessible to our observation, but the question of the existence of such powers is essentially outside the limits of physical investigation.

The reduction here effected of the relation between mind and matter to a relation between mind and energy does not explain the nature of the relation, which still remains an apparently inscrutable mystery, for we can at present form no manner of conception of the nature of the process by which mind effects, or becomes sensible of, changes in the distribution of energy. What has been proved is that the observed effects on the material world, of the conscious and unconscious activities of living organisms, are expressible in terms of changes in the distribution of energy, and that the power of effecting similar changes upon a sufficiently extended scale is all that is required to account for the formation of the material universe from any initial distribution of its total energy, and for the establishment of all the known laws or conditions under which it now exists. The term mind has been employed to designate the basis of observed activity in the case of living organisms, while the basis of the extended activity required to account for the material order of nature has been called the universal mind. Since the concept of the universal mind contains no characteristics differing in kind from those observed in the case of man and other living organisms, it is a purely scientific concept arrived at by legitimate inference from observed facts, and it has been proved adequate for the explanation of the observed phenomena. I shall therefore consider it no longer as a working hypothesis, but as an established scientific theory which can only be called in question by the presentation of an alternative scientific theory equally capable of accounting for the observed facts. No such theory has yet been propounded.

I shall now proceed to show that the theory of the universal mind accounts for the origin of life, with respect to which Lord Kelvin declared (*Nineteenth Century and After,* June 1903):—

> "I cannot say, with regard to the origin of life, science neither affirms nor denies creative power. Science positively affirms creating and directive power, which she compels us to accept as an article of belief."

Innumerable attempts have been made to demonstrate the possibility of *Abiogenesis*, or the spontaneous transformation of non-living into living substance, but every one has been a complete failure, the only apparent successes having been proved by later and more exact investigation to have been due to imperfections in the experimental methods preventing the complete ex-

clusion of living germs from the apparatus employed. Tyndall wrote in 1878:—

"I affirm that no trustworthy shred of experimental testimony exists to prove that life in our day has ever appeared independently of antecedent life."[1]

His statement is still as indisputable as when it was made, and the theory of *Biogenesis*, or the derivation of life only from antecedent life, is now, as then, the only one supported by experi mental evidence. Our present point of view does not, however, in any way suggest the impossibility, or even the improbability, of the discovery of the conditions under which the inception of life occurs, and such a discovery would in no way affect the argument. Many biologists have expressed the opinion that Abiogenesis *must* have occurred at some earlier period of the earth's history, but not one has produced a single shred of evidence in support of his opinion, and such unsupported opinions are of no scientific value whatever. There exists, therefore, as far as experimental evidence can prove it, an absolute wall of demarcation between non-living and living substance. This does not consist in any mechanical or molecular difference in the nature of the material structure of living organisms from that of inorganic matter, for all the laws of the mechanical and molecular schemes are, as we shall presently see, absolutely continuous throughout all kinds of matter, non-living and living. We shall see that the distinction consists in the advent of what I have called mind, and it will be shown that a rudimentary power of adaptation to the environment is a characteristic of even the lowest forms of life, and is exhibited to an increasing extent in the higher forms, both vegetable and animal. Moreover, every living organism possesses the power of reproducing forms of life similar to its own; that is to say, it has the power of transmitting mind to, or, in other words, conferring the gift of life upon, previously inanimate matter. This power is characteristic of living substance, that is to say, substance consisting of matter associated with mind. We have, therefore, to add this observed power of living organisms to those included in the concept of the universal mind, which is therefore the one lord and giver of life.

J. C. Bose has shown, in an extensive series of researches, that response to electrical excitation affords the most definite and reliable criterion of an animal or plant being living or dead. A plant, for example, which has been exposed for a short period to an unduly low temperature, may, for the moment, show no outward signs of injury, but if it refuse to respond to electrical excitation,

[1] *Nineteenth Century*, 1878, p. 507.

it will certainly begin to wither very shortly, and is already past all hope of revival; it is, in fact, dead. Electrical excitation in plants gives rise to response curves exactly resembling those obtained from the muscles of animals, when subjected to mechanical strain, and similar curves can be obtained from inorganic substances, such as metallic powders subjected to molecular strain, by the action of electric radiation, the reason being that the existence of molecular strain in inorganic substances is shown by changes in electric conductivity. The response curves are obtained by taking the galvanometer deflections as ordinates, and the times of exposure, or recovery, as abscissæ. One of the most suitable substances for the observations from which these curves are obtained is found to be slightly warmed magnetic oxide of iron, as the effects, in the case of this substance, are produced so slowly as to enable them to be recorded without difficulty, while the recovery, under the right temperature conditions, is rapid, and there are no secondary chemical reactions. There is always a latent period between the stimulus and the response, which latter rises to a maximum after the cessation of the stimulus, followed by a rapid recovery. With some substances a tendency to after-oscillation can be observed, but this is never so noticeable as in the muscular strain curves. When a second stimulus is applied, following one of moderate amount, there is a summation of the effects; but if the second stimulus follows one strong enough to produce the maximum effect, there is no further result. When stimuli follow with slow intermission, a broken curve is obtained, exactly similar to the corresponding muscular ones; while, with rapid intermissions, the curve becomes continuous, as in the tetanic state, or *rigor*, of a muscle. If electrical response could be taken as the sole criterion of life, positively as well as negatively, then we should be justified in saying that a metal can be killed, as shown by its ceasing to give a response curve, or that it can be drugged, or partially poisoned, by exposure to solutions of deleterious substances. In the latter cases the response is greatly weakened; but, until it has completely disappeared, the metal may be restored to its normal state of health by the administration of an antidote. When the response has entirely disappeared this is no longer possible. Here, however, the analogy between metals, and the *protoplasm*, or *physical basis of life*, of an animal or plant, breaks down. The dead metal can be restored to its normal state by various physical and chemical, that is to say, molecular, processes, while in the case of the protoplasm, this can only be effected by its absorption into the substance of a living animal or plant. The analogy between metals and organic forms may be further illustrated by the following quotation from the opening chapter of Floris Osmond's

Microscopic Analysis of Metals, translated and edited by T. E. Stead :—

"Metals are neither simple bodies nor are they inert. The ultimate analysis of an alloy does not tell us how the elements present are combined among themselves, or whether they form isolated definite compounds or solutions, homogeneous or not. The proximate analysis, which, in addition, is not often possible in the present state of chemistry, does not tell us how the constituents, supposed to be chemically separated, are individually organised or topographically distributed, as compared one with another. Histological organisation and anatomical distribution, supposing they are known, do not tell us what modifications the actual condition would undergo under the influence of certain changes of temperature or pressure. Finally, these normal modifications, supposing them to be known in their turn, might be considerably disturbed by an infinitesimal quantity of some foreign impurity, some times inevitable, and the very presence of which is, at times, unsuspected.

"Thus we are naturally led to establish in the study of metals several subdivisions quite analogous to those of medical science, and to speak of metallography as anatomical (histologic), biological, and pathological.

"The first distinguishes and defines the different constituents of which an alloy may consist, by observing their optical characteristics, their chemical characteristics, or their mechanical characteristics, and describes their forms, crystalline or otherwise, measures their absolute or relative dimensions, and examines the planes of weakness which separate one from the other or traverse each of them separately.

"The second subdivision is occupied with determining how the composition, form, dimensions, and relations of the different constituents, determined in a fixed state in a given sample, are connected with possible conditions of calorific or mechanical treatment which the alloy under consideration may undergo during the process of its manufacture or employment.

"The third subdivision deals with the influence of errors of treatment, and of the presence of impurities which constitute, as it were, a particular diathesis. At the same time, when effects have once been connected with their causes by a preparatory study, it often allows us to group them together again for the solution of the problems given by practical work."

Another remarkable analogy between organic and inorganic forms is exhibited by crystals, as may be illustrated by a quotation from R. A. Hadfield's Presidential Address to the Iron and Steel Institute in 1905, in which he observes :—

"The self-repairing power of a crystal is like that of a human being. A slightly abraided crystal of alum, dipped for an instant into a saturated aqueous solution of that salt, completely repairs the injured parts ; and if the solution contain other salts dissolved in it, the crystal repairs itself with alum only, refusing to unite with the other substances, provided they do not belong to the same crystalline system. This selective power is like that of a bone, muscle, or nerve, each of which will only take to itself from the blood its own proper ingredients."

Innumerable examples might be given of the simulation of vegetable forms by the accretion of inorganic particles, but it will be sufficient to adduce the one familiar to all, the delicate, fern-like tracery so often produced by frost on the moisture of a window pane. Even the colloid, or jelly-like, structure of protoplasm itself finds its inorganic counterpart in the hydrofluosilicates. It will be unnecessary further to illustrate what all physiological research confirms—the absolute continuity of all known mechanical and molecular laws throughout living, equally with non-living, matter. We may therefore proceed to consider the next point, the appearance and development of purposive action simultaneously with the appearance and development of living organisms.

Some biologists have suggested that something which we may call *mind-stuff*, to use a term of Professor Clifford's, is secreted, that is to say, developed, in the brain cells of men and animals. This does not, however, account for the facts, even from the most restricted and purely biological point of view, for it does not explain the rudimentary purposiveness observed in the earliest and least developed forms of plant and animal life. Other biologists have endeavoured to meet this difficulty by assuming mind-stuff to be formed in every living cell, and Haeckel calls this supposed secretion the cellular, or atomic, soul. Even in this form, however, the hypothesis does not account for the biological facts. That most lowly of all recognised living organisms—indeed, we can scarcely call it an organism—the Moneron, already shows signs of purposive action, shows that something of the nature of mind-stuff must be *associated* with, which does not necessarily mean developed out of, the single cell of which it consists. Haeckel, who has made an exhaustive study of this early form of life, tells us that in its simplest, and therefore, probably, its earliest form, the whole creature consists of a single cell without either nucleus or integument, simpler therefore than any of the constituent cells of higher organisms, a mere globular mass of slime.[1] Yet even so lowly a creature can be actuated by the urge of hunger. I do not say it can feel the urge, for we cannot conceive of anything

[1] Sir Edwin Ray Lankester (*Kingdom of Man*, p. 118) tells us that he is inclined to agree with the conclusion, arrived at by some of the latest workers in histology, that the evidence for the existence, at the present time, of so simple an organism is insufficient, and that the simplest organism now in existence is a nucleated cell. If he were correct, the argument in the text would be unaffected, though the illustration would not be quite so striking. Haeckel, however, in his most recent work, *The Wonders of Life*, p. 23, maintains that these histologists are wrong, and I understand that on this point his authority is unquestionably a high one, as his observations on the Monera have been extensive and detailed.

approaching even the elementary form of consciousness which we call sensation in so undeveloped a form of life; we cannot call it either animal or plant. Haeckel takes it as the starting point from which these two forms of life diverge. When impelled by hunger, it throws out tentacle-like protuberances, and when these come in contact with a particle of food (they are able to distinguish suitable from unsuitable particles), smaller tentacles are protruded from the sides of the larger ones so as to enlace the particle in a network of slime, which draws it into the interior, where it becomes gradually absorbed by a process of solution, and as soon as it is thus drawn in, the creature resumes its globular shape.

If a rudimentary form of purposive action is thus exhibited at the very starting point of plant and animal life, we should expect to find some indications of it in the higher forms of plants, as well as in the higher animals. It has usually been assumed that nothing of the kind is observable in plant life, but I believe that some, at least, of our leading botanists are now being driven to the conclusion that purposive action is to be found in plants as well as animals. I cannot myself conceive how else we can explain the manner in which the roots of a plant will seek out and find spots where suitable food or water are present. The roots seem to travel directly towards it, sometimes for distances of many yards, just as though the plant knew by some subtle instinct the exact spot where it was to be found. I might give many more instances, but that my own biological knowledge is too limited to be worth calling knowledge, and therefore my personal opinions on this matter can have no possible weight. I prefer, therefore, to rely upon the statements of acknowledged authorities in reference to biological facts, and shall confine my further illustrations of the purposive actions of plants to statements of the greatest of all biological authorities, Charles Darwin, beginning with some extracts from the "Concluding Remarks" in his work on *Climbing Plants*, p. 197.

"If we inquire how a petiole,[1] a branch or flower-peduncle,[2] first became sensitive to a touch, and acquired the power of bending towards the touched side, we get no certain answer. Nevertheless, an observation by Hofmeister well deserves attention, namely, that the shoots and leaves of all plants, whilst young, move after being shaken. Kerner also finds, as we have seen, that the flower-peduncles of a large number of plants, if shaken or gently rubbed, bend to this side. And it is young petioles and tendrils, whatever their homological nature may be, which move on being touched. It thus appears that climbing plants

[1] A leaf stalk. [2] A flower stalk.

have utilised and perfected a widely distributed and incipient capacity, which capacity, as far as we can see, is of no service to ordinary plants. If we further inquire how the stems, petioles, tendrils, and flower-peduncles of climbing plants first acquired their power of spontaneously revolving, or, to speak more accurately, of successively bending to all points of the compass, we are again silenced, or at most can only remark that the power of moving, both spontaneously and from various stimulants, is far more common with plants than is generally supposed to be the case by those who have not attended to the subject."

Again, p. 202 :—

"The most interesting point in the natural history of climbing plants is the various kinds of movement which they display in manifest relation to their wants. The most different organs—stems, branches, flower-peduncles, petioles, mid-ribs of the leaf and leaflets, and apparently aerial roots—all possess this power. The first action of a tendril is to place itself in a proper position. For instance, the tendril of Cobæa first rises vertically up, with its branches divergent and with the terminal hooks turned outwards; the young shoot at the extremity of the stem is at the same time bent to one side, so as to be out of the way. The young leaves of clematis, on the other hand, prepare for action by temporarily curving themselves downwards so as to serve as grapnels.

"Secondly, if a twining plant or a tendril gets by any accident into an inclined position, it soon bends upwards, though secluded from the light. The guiding stimulus no doubt is the attraction of gravity, as Andrew Knight showed to be the case with germinating plants. If a shoot of any ordinary plant be placed in an inclined position in a glass of water in the dark, the extremity will, in a few hours, bend upwards; and if the position of the shoot be then reversed, the downward-bent shoot reverses its curvature; but if the stolon of a strawberry, which has no tendency to grow upwards, be thus treated, it will curve downwards in the direction of, instead of in opposition to, the force of gravity. As with the strawberry, so it is generally with the twining shoots of the *Hibbertia dentata*, which climbs laterally from bush to bush; for these shoots, if placed in a position inclined downwards, show little and sometimes no tendency to curve upwards.

"Thirdly, climbing plants, like other plants, bend towards the light by a movement closely analogous to the incurvation which causes them to revolve, so that their revolving movement is often accelerated or retarded in travelling to or from the light. On the other hand, in a few instances tendrils bend towards the dark.

"Fourthly, we have the spontaneous revolving movement which is independent of any outward stimulus, but is contingent on the growth of the part, and on vigorous health; and this again of course depends on a proper temperature and other favourable conditions of life.

"Fifthly, tendrils, whatever their homological nature may be, and the petioles or tips of the leaves of leaf-climbers, and apparently certain roots, all have the power of movement when touched, and bend quickly towards the touched side. Extremely light pressure often suffices. If

the pressure be not permanent, the part in question straightens itself and is again ready to bend on being touched.

"Sixthly, and lastly, tendrils, soon after clasping a support, but not after a mere temporary curvature, contract spirally. If they have not come into contact with any object, they ultimately contract spirally, after ceasing to revolve; but in this case the movement is useless, and occurs only after a considerable lapse of time.

"With respect to the means by which the various movements are effected, there can be little doubt, from the researches of Sachs and H. de Vries, that they are due to unequal growth; but from the reasons already assigned, I cannot believe that this explanation applies to the rapid movements from a delicate touch."

Again, p. 206:—

"It has often been vaguely asserted that plants are distinguished from animals by not having the power of movement. It should rather be said that plants acquire and display this power only when it is of some advantage to them; this being of comparatively rare occurrence, as they are affixed to the ground, and food is brought to them by the air and rain. We see how high in the scale of organisation a plant may rise when we look at one of the more perfect tendril-bearers. It first places its tendrils ready for action, as a polypus places its tentacula. If the tendril be displaced, it is acted on by the force of gravity and rights itself. It is acted on by the light, and bends towards or from it, or disregards it, whichever may be most advantageous. During several days the tendrils or internodes, or both, spontaneously revolve with a steady motion. The tendril strikes some object, and quickly curls round and firmly grasps it. In the course of some hours it contracts into a spire, dragging up the stem, and forming an excellent spring. All movements now cease. By growth the tissues soon become wonderfully strong and durable. The tendril has done its work, and has done it in an admirable manner."

My next illustrations will be taken from Darwin's observations on the common Sun-Dew, *Drosera rotundifolia*; and *Dionæa muscipula*, commonly known as Venus's fly-trap; as described in his work on *Insectivorous Plants*.

The Sun-Dew is a tiny plant bearing from two or three, to five or six, leaves having the whole of their upper surfaces covered with minute, gland-bearing, hair-like filaments, or tentacles, as Darwin calls them, owing to their manner of acting. If a small organic or inorganic object be placed on the glands in the centre of a leaf, if a fly alight on them, for example, a motor impulse is transmitted to the marginal tentacles. The nearer ones are first affected and slowly bend towards the centre, and then those farther off, until at last all become closely inflected over the object, which, if a living one, is prevented from escaping, by means of the viscid substance which extrudes from the glands, the process occupying from one to five or more hours, according

to the size and nature of the object. The bending of the tentacles takes place entirely from the base.

Darwin observes, p. 276 :—

"If the bending place of a tentacle receives an impulse from its own gland, the movement is always towards the centre of the leaf; and so it is with all the tentacles, when their glands are excited by immersion in a proper fluid. [The effects of immersing the leaf in various fluids are described in detail in the volume.] The short ones in the middle part of the disc must be excepted, as these do not bend at all when thus excited. On the other hand, when the motor impulse comes from one side of the disc, all bend with precision towards the point of excitement, wherever this may be seated. This is in every way a remarkable phenomenon, for the leaf falsely appears as if endowed with the senses of an animal. It is all the more remarkable as, when the motor impulse strikes the base of a tentacle obliquely with respect to its flattened surface, the contraction of the cells must be confined to one, two, or a very few rows at one end. And different sides of the surrounding tentacles must be acted on, in order that all should bend with precision to the point of excitement.

"The motor impulse, as it spreads from one or more glands across the disc, enters the bases of the surrounding tentacles, and immediately acts on the bending place. It does not, in the first place, proceed up the tentacles to the glands, exciting them to reflect back an impulse to their bases. Nevertheless, some influence is sent up to the glands, as their secretion is soon increased and rendered acid; and then the glands, being thus excited, send back some other influence (not dependent on increased secretion, nor on the inflection of the tentacles), causing the protoplasm to aggregate in cell beneath cell. [It had been previously shown that after excitation the cells in the tentacles, instead of being filled with a homogeneous purple fluid, as before, now contain variously shaped masses of purple matter, suspended in a more or less colourless fluid.] This may be called a reflex action, though probably very different from that proceeding from the nerve-ganglion of an animal; and it is the only known case of reflex action in the vegetable kingdom."

Another insectivorous plant, Venus's fly-trap, is rightly designated by Darwin as one of the most wonderful plants in the world. Each leaf consists of two lobes attached to a foliaceous petiole or foot-stalk. To continue in Darwin's words, p. 287 :—

"The two lobes stand at rather less than a right angle to each other. Three minute-pointed processes or filaments, placed triangularly, project from the upper surfaces of both; but I have seen two leaves with four filaments on each side, and another with only two. These filaments are remarkable for their extreme sensitiveness to a touch, as shown not by their own movement, but by that of the lobes. The margins of the leaf are prolonged into sharp rigid projections which I will call spikes, into each of which a bundle of spiral vessels enters. The spikes stand in such a position that, when the lobes close, they interlock like the teeth of a rat-trap. The mid-rib of the leaf, on the lower side, is strongly developed and prominent. The upper surface of the leaf is

thickly covered, excepting towards the margins, with minute glands of a reddish or purplish colour, the rest of the leaf being green. There are no glands on the spikes, or on the foliaceous foot-stalk. The glands are formed of from twenty to thirty polygonal cells filled with purple fluid."

The filaments from their tips to their bases are extremely sensitive to a momentary touch by any solid body, and when so touched the trap instantly closes. In spite of this sensitiveness, no effect is produced by a fine jet of air or water on the filaments. The trap will therefore not be sprung by wind or rain. When an insect of sufficient size to satisfy the plant is caught, the lobes remain closed for a period of many days, until the insect is digested by the aid of secretions exuded from the glands. If the plant is tricked by means of small objects unsuitable for food being dropped on to the sensitive filaments, no secretion is exuded from the glands, and after some twenty-four hours or so, the lobes open and allow them to drop out. Moreover, if the trap is sprung by an insect too small to be suitable, it is able to escape between the interlocked spikes, and, after an interval of some twenty-four hours, the lobes open again, so that the trap is reset.

If small particles of slightly damp nitrogenous substances suitable for food are laid upon the lobes, without touching the sensitive filaments, the lobes close slowly and remain closed until the food is digested. Other objects, such as wood, cork, moss, paper, stone or glass, may be left for any length of time on the surface of a leaf. The lobes do not close, and no secretion is exuded. The plant simply ignores them. The level of purposiveness is higher than that attained by the Sun-Dew.

The association of mind with so rudimentary an organism as the Moneron in its simplest form, consisting of a mere drop of the jelly-like substance known as protoplasm, a drop of primal slime as we may call it, is entirely unaccountable on the basis of any theory of development of mind as the result of organisation. Mind here appears as antecedent to any perceptible organisation. This is, however, in exact accordance with the theory that this mind is the first rudimentary differentiation of the universal mind, and instead of being the result of, is the active agent in, the development of the organisation of which this rudimentary form appears to represent the starting point. According to the *transformist*, or development, theory of the origin of all living forms, of which Lamarck, Darwin, and Wallace have been the ablest exponents, the innumerable forms of living things now existing on the surface of the globe, or which have existed during past ages, have all arisen from one or more rudimentary organisms of which the Moneron may be taken as a type, if it does not form the actual starting point. It cannot be maintained that this theory of the

continuous evolution of living forms has ever yet been completely established upon an irrefutable observational basis. When primal intelligence is excluded from consideration, the theory simply bristles with apparently insuperable difficulties, even when the first step, the advent of life, is taken for granted, without attempting to account for it. It cannot be said that all the difficulties entirely disappear with the admission of primal intelligence, but they no longer appear insuperable.

The present practically universal acceptance of the theory of evolution by scientific thinkers is not to be ascribed to its complete demonstration on the basis of observed facts, but to its being the only theory of the origin of living beings which at all accords with our experience of the ordered processes of nature. It is the only theory at present conceivable which makes scientific investigation of that origin a possibility. Professor S. H. Vines, in his Presidential address to the Linnæan Society in 1902, thus expressed himself in reference to our knowledge of the genealogy of organic life:—

> "Though here and there fragments of the mosaic seem to have been successfully pieced together, the main outlines, even, of the great picture are as yet but dimly discernible. The fact that organic evolution should have proceeded so far as it has within such limits of time as may reasonably be allowed, admits, to my mind, of no other interpretation than that variation is not indeterminate, but, as Lamarck and Nägeli have urged, there must exist in living matter a certain inherent tendency or bias in favour of variation in the higher direction. It is this tendency or bias that I venture to regard as the primordial factor."

Natural selection, or the survival of the fittest, which was propounded by Darwin and Wallace as an explanation of the means by which natural variations favourable to the organism in its existing environment, which may differ from the environment of its progenitors, are perpetuated, is in all probability, after the agency from which these variations arise, the most potent, although not the sole, further agency involved. Wallace has shown [1] that constant variations are observed in all the parts of organisms derived from the same parents, and that they frequently amount to as much as 20 per cent. of the part implicated. He considers that these variations, perpetuated by natural selection, are quite sufficient to account for the variability of species essential to a theory of evolution. A long series of very successful attempts to develop permanent and valuable varieties of flowering plants led de Vries to a somewhat different conclusion. He found that most of these constant variations were not strictly hereditary, so that it was not in general possible to obtain permanent varieties by their continued artificial selection.

[1] *Darwinism*, chapter iii.

He found, however, that from time to time variations of much greater extent made their appearance, quite suddenly, and were in general hereditarily permanent. These hereditary variations, giving rise to permanent varieties, or *elementary species*, he called *Mutations*. The mutation theory is not accepted by Wallace, but it has received remarkable confirmation in the experiments carried on during the last twenty-three years by H. Nillson at the seed culture institute at Svalöf, in Sweden, which was founded "to obtain, by systematic selection, new and better sorts of agricultural plants that would be of finer quality and of higher value than any that could be obtained in the market."[1] The mutation hypothesis certainly appears to remove one of the greatest difficulties of the transformist theory. Natural selection can take no account of the interests of the future progeny. It merely effects the elimination of the forms less suitable to the existing environment, and cannot, therefore, promote changes in a direction which would be ultimately favourable, but not immediately beneficial; and it is difficult to see how many of the transitions, such, for example, as that from a reptile to a bird, could be effected by continuous variations of such a character as to be beneficial to the developing organism at every stage of its development. And these are the only variations which could be perpetuated by natural selection. Moreover, no explanation on the basis of the mechanical and molecular schemes, usually designated by biologists as a "mechanical" explanation, has ever been given of the origin of the observed variations in the progeny of the same parents, which form the essential foundation of the transformist theory. Whether the mutation hypothesis be confirmed or disproved, that is to say, whether the ordinary constant variations are or are not hereditary, the occurrence of occasional variations of exceptional extent has been definitely proved by the observations both of de Vries and Nillson, and if these are regulated in frequency and extent by the purposive intelligence of the universal mind, acting partially or entirely through the individual mind of the organism, the whole difficulty of accounting for these transitions disappears absolutely.

That the mind of an organism should be capable of effecting these variations in the progeny developed from its own mentality and bodily structure is perfectly explicable when it is remembered that they are simply variations in mind and energy distribution, corresponding to the several units constituting the progeny. That it actually possesses this capacity is demonstrated by direct observation in the case of man and the higher animals. The possibly deleterious effects on the bodily health and structure,

[1] A. A. W. Hubrecht, *Contemporary Review*, June 1909, p. 738.

as well as on the mentality of the progeny, of subjecting the gravid female to emotional disturbance, is universally known and acted upon in the case of the human female. The effect on the bodily health and structure is, moreover, well known to every observant breeder of animals. The same influence of the parent mind affords the simplest explanation that has yet been suggested of the building up of the embryo, from the initial protoplasmic germ, into a form which will ultimately resemble that of the parent. The protoplasm of a mushroom is physically indistinguishable from that of a man, but, as Sir William Thistleton Dyer observes:—[1]

"It is no doubt true that a particle of fungoid differs in no appreciable physical respect from one of human protoplasm, yet the former will never emerge from the fate of the humble mushroom, while the other may be instinct with the thoughts of a Prime Minister."

If the substances are physically indistinguishable, then, unless the observed distinctions in the results of growth are to be attributed, wholly or partially, to differences in minutiæ of atomic dimensions, below the reach of observation, we can only suppose them to be determined entirely by the nature of the mind associated with, and acting upon, either germ. Huxley, describing the development under the microscope of the germ of a frog, observes:—[2]

"Strange possibilities lie dormant in that semi-fluid globule. Let a moderate supply of warmth reach its watery cradle and the plastic matter undergoes changes so rapid, and yet so steady and purpose-like in their succession, that one can only compare them to those operated by a skilled modeller upon a formless lump of clay. As with an invisible trowel the mass is divided and subdivided into smaller and smaller portions, until it is reduced to an aggregation of granules not too large to build withal the finest fabrics of the nascent organism. And, then, it is as if a delicate finger traced out the line to be occupied by the spinal column, and moulded the contour of the body; pinching up the head at one end, the tail at the other, and fashioning flank and limb into due proportions in so artistic a way that, after watching the process hour by hour, one is almost involuntarily possessed by the notion that some more subtle aid to vision than an achromatic would show the hidden artist, with his plan before him, striving with skilful manipulation to perfect his work."

In this case the modeller can be none other than the mind already transmitted from the parents to the egg, acting as a differentiated portion of the all-pervading universal mind, yet never severed from its influence.

[1] *Nature*, June 5, 1902, p. 121.
[2] *Lay Sermons*, 6th edition, p. 261.

A very interesting point in the embryonic development of organic forms, and one which affords a most cogent argument in favour of the theory of organic evolution, is that the embryos of all animals and plants are found to pass in a few days, or weeks, or months, as the case may be, through a succession of stages corresponding to the whole presumed course of ancestral evolution, extending over a period of millions of years, and yet they always terminate in a product of similar type to the parents.

Professor Huxley tells us[1] that the yolk of the egg first

"undergoes division or segmentation, as it is called; the ultimate products of that segmentation constitute the building materials for the body of the young animal; and this is built up round a primitive groove, in the floor of which a notochord is developed. Furthermore, there is a period in which the young of all these animals resemble one another, not merely in outward form, but in all essentials of structure, so closely, that the differences between them are inconsiderable, while in their subsequent course they diverge more and more widely from one another. And it is a general law that the more closely any animals resemble one another in adult structure, the longer and the more intimately do their embryos resemble one another; so that, for example, the embryos of a snake and of a lizard remain like one another longer than do those of a snake and a bird; and the embryos of a dog and of a cat remain like one another for a far longer period than do those of a dog and a bird, or of a dog and an opossum, or even than those of a dog and a monkey."

And again:—

"It is very long before the body of the young human being can be readily discriminated from that of the young puppy; but at a tolerably early period the two become distinguishable by the different forms of their adjuncts, the yolk-sac and the allantois." . . . "But exactly in those respects in which the developing man differs from the dog, he resembles the ape." . . . "So that it is only quite in the latter stages of development that the young human being presents marked differences from the young ape, while the latter departs as much from the dog in its development as the man does. Startling as this last assertion may appear to be, it is demonstrably true, and it alone appears to me sufficient to place beyond all doubt the structural unity of man with the rest of the animal world, and more particularly and closely with the apes."

Wallace[2] enumerates a few of the curious details in which man passes through stages common to the lower animals, viz.:—

"At one stage the os coccyx projects like a true tail, extending considerably beyond the rudimentary legs. In the seventh month the convolutions of the brain resemble those of an adult baboon. The great

---

[1] *Man's Place in Nature*, p. 64.
[2] *Darwinism*, p. 449.

toe, so characteristic of man, forming the fulcrum which most assists him in standing erect, in an early stage of the embryo is much shorter than the other toes, and instead of being parallel with them, projects at an angle from the side of the foot, thus corresponding with the permanent condition in the Quadrumana. Numerous other examples might be quoted, all illustrating the same general law."

We have seen how completely satisfactory an explanation is afforded by the theory of the universal mind of the absolute continuity in mechanical and molecular laws throughout living as well as non-living matter, coupled with the absolute line of demarcation between them which is marked by the advent of life. A somewhat analogous phenomenon is presented by the continuity observed in the structure and development of all living organisms from the Moneron to man, together with the line of demarcation between man and all other organisms which is marked by the exclusively human faculty of evolving articulate speech as a medium for the expression and growth of reason. The evidence collected by Romanes, and published in his work on *Animal Intelligence*, shows the existence of rudimentary reasoning power in the minds of the higher animals. The line of demarcation is, therefore, much less distinctly drawn than that between non-living and living matter, but that there is such a demarcation would appear to follow as a necessary consequence of the observed facts that no animal has ever been found to possess the faculty of expressing itself in terms of language, and that this faculty is, on the other hand, possessed even by the lowest existing races of men. In reference to this point, Professor Max Müller, in his Presidential address to the Anthropological Section of the British Association in 1889, expressed himself as follows:—

"What does Bunsen consider the real barrier between man and beast? It is language, which is unattainable, or at least unattained, by any animal except man. You know how for a time, and chiefly owing to Darwin's predominating influence, every conceivable effort was made to reduce the distance which language places between man and beast, and to treat language as a vanishing line in the mental evolution of animal and man. It required some courage at times to stand up against the authority of Darwin, but at present all serious thinkers agree, I believe, with Bunsen, that no animal has ever developed what we mean by rational language, as distinct from mere utterances of pleasure or pain, a subject lately treated with great fulness by Professor Romanes. Still, if all true science is based on facts, the fact remains that no animal has ever found what we mean by a language; and we are fully justified, therefore, in holding with Bunsen and Humboldt, as against Darwin and Romanes, that there *is* a specific difference between the human animal and all other animals, and that that difference consists in language as the outward manifestation of what the Greeks meant by *Logos.*"

Wallace,[1] while strongly insisting on these two lines of demarcation, introduces another, intermediate, stage between them, the advent of consciousness. After dealing with the first stage, "the change from inorganic to organic," he writes :—

"The next stage is still more marvellous, still more completely beyond all possibility of explanation by matter, its laws and forces. It is the introduction of sensation or consciousness, constituting the fundamental distinction between the animal and vegetable kingdoms. Here all idea of mere complication of structure producing the result is out of the question. We feel it to be altogether preposterous to assume that at a certain stage of complexity of atomic constitution, and as a necessary result of that complexity alone, an *ego* should start into existence, a thing that *feels*, that is *conscious* of its own existence. Here we have the certainty that something new has arisen, a being whose nascent consciousness has gone on increasing in power and definiteness till it has culminated in the higher animals. No verbal explanation or attempt at explanation—such as the statement that life is the result of the molecular forces of the protoplasm, or that the whole existing organic universe from the amœba up to man was latent in the fire-mist from which the solar system was developed—can afford any mental satisfaction, or help us in any way to a solution of the mystery."

I am in complete agreement with the principle of this argument, except that instead of identifying sensation and consciousness, it appears to me that the observed facts indicate a gradual ascent from the most rudimentary form of sensation, accompanying the first advent of life, to the full consciousness of an *ego* knowing its own existence; and that this consciousness of an *ego* or *self*, even in its most rudimentary form, appears first in man, and in its fully developed form, only in man after he has attained a high grade of development. Wallace's second stage I therefore regard as spread over the whole interval between the first line of demarcation, the advent of life, and the second line of demarcation, the advent of man.

Previously to the advent of life the universal mind is manifested only as a formative and directive power not visibly in direct association with matter. At this point mind first becomes so closely associated with matter that their properties are distinguishable only through the knowledge obtained by the comparison of observations on non-living and living matter. If living matter alone were available for observation, a purely pantheistic scheme would be perfectly adequate to the expression of the observed phenomena of nature, although philosophically inadequate, as identifying the Absolute and Unconditioned with its phenomenal manifestations.

[1] *Darwinism*, pp. 474–475.

The observed distinction between non-living and living matter shows it to be an inadequate representation even of observed phenomena. If it were possible to take the further step of demonstrating the derivation of energy from the universal mind, the latter would then form a complete representation of all observed phenomena, but even then it would be unphilosophical to identify such a phenomenal representation with the Absolute. To do this would be to limit the Absolute to its manifestations relatively to ourselves, and although these manifestations are all that can possibly become known to us, we should still be committing the absurdity of imposing conditions upon the Unconditioned. Drummond adduces reasons for supposing that the ascent of the material organism on our globe has already attained its culmination in the present bodily structure of man,[1] and that the continuance of evolution is to be looked for in the future mainly in the development of his higher mental and spiritual faculties. In this process the influence of natural selection is being gradually displaced by the purposive directive intelligence of man, and the analogy of the past would lead us to anticipate a steadily increasing influx of these higher faculties from the all-pervading universal mind as his upward progress makes him capable of their assimilation. The ethical faculty, in the sense of the faculty by which we endeavour to determine the value or worth of conduct and the truth of judgments passed upon it, cannot be accounted for by a "naturalistic scheme," as it is often called; that is to say, a scheme according to which, in the words of Professor W. R. Sorley,[2] "the completest account of the world as a whole which is possible is held to be the description of it in physical[3] terms: the spiritual factor is held to be dependent, if not illusory." It would be impossible to justify this statement in a few pages, but it is demonstrated with the most convincing logic and lucidity in the work referred to. Huxley takes an extreme, and not, in my opinion, a justifiable, position in support of this view. He maintains that, as Professor Sorley[4] has expressed it elsewhere,

"the cosmic order has nothing to say to the moral order, except that, somehow or other, it has given it birth; the moral order has nothing to say to the cosmic order, except that it is certainly bad."

---

[1] See *The Ascent of Man*, H. Drummond, chapter iii.

[2] *Ethics of Naturalism*, p. 17.

[3] Professor Sorley evidently employs the term "physical" in the sense of material, and not with the wider signification which I have given to it, as including mental, as well as molecular and chemical, phenomena.

[4] *Recent Tendencies in Ethics*, p. 47.

To quote Huxley's own words,[1]

"the practice of that which is ethically best, what we call goodness or virtue, involves a course of conduct which, in all aspects, is opposed to success in the cosmic struggle for existence. . . . The ethical progress of society depends, not on imitating the cosmic process, still less in running away from it, but in combating it."

In the earlier phases of life the egoistic principle, the struggle for life, is certainly the dominant factor; but the altruistic principle, or the struggle for the life of others, as Drummond calls it, makes its appearance at an early stage in the phenomena of reproduction. As Herbert Spencer observes :—[2]

"What is the ethical aspect of these [altruistic] principles? In the first place, animal life of all but the lowest kinds has been maintained by virtue of them. Excluding the *Protozoa*, among which their operation is scarcely discernible, we see that without *gratis* benefits to offspring, and earned benefits to adults, life could not have continued. In the second place, by virtue of them life has gradually evolved into higher forms. By care of offspring, which has become greater with advancing organisation, and by survival of the fittest in the competition among adults, which has become more habitual with advancing organisation, superiority has been perpetually fostered and further advances caused."

Given the existence of this altruistic care for offspring, even in the most rudimentary form, and its variability in different individuals, natural selection is perfectly competent to account for its maintenance and development. Both its original existence and its variability are explicable from the point of view I have advocated, but have never been accounted for on any so-called mechanical principle. When evolution comes under the influence of man's purposive intelligence and foresight, natural selection gradually gives way to artificial, or purposive, selection, and this is pre-eminently true of the higher stages of ethical evolution. As Professor Sorley tells us :—[3]

"It must be borne in mind that in strictness 'natural selection' is not selection at all: it is only a negative, not a positive, process. The positive tendency comes from another source altogether, from a tendency within the organism. On the lower levels of life its chief manifestation is the unexplained tendency to variation. In man the force is reflective and rational: ends are anticipated in idea and deliberately pursued; and, in so far as the activity is wisely directed, the negative process of selection called 'natural' is displaced by a positive selection which is intelligent. The course of human development is altogether misinter-

[1] *Evolution and Ethics*, pp. 81–83.
[2] *Principles of Ethics*, vol. ii. p. 5.
[3] *The Ethics of Naturalism*, pp. 318–320.

preted if we overlook either the operation of subjective selection on the part of the individual who strives to accomplish his end, or the organised operation of the same force as it is exhibited in social selection. To the latter natural selection bears some analogy in the results it produces; but they are fundamentally distinct in their nature and mode of working. Evolution works in many ways, and natural selection is neither its first nor its final method. So far, therefore, as we are concerned with the questions of fact and history commonly classed as ethical—that is to say, with the nature and growth of moral institutions and moral ideas—the assertion of moral development does not imply that the process of evolution has been determined by the same method as that which rules in biology, any more than it implies that it has been determined by the same method as that which rules astronomical or cosmical changes. The course of moral development may begin at a period when natural selection is in the ascendant, but it rises out of it to a higher stage in which the influence of natural selection wanes and reason dominates the process. The facts of the moral life—whether they are of the nature of social institution or whether they belong to the inner world of ideas—cannot be explained by natural selection alone. To understand them we must recognise the obvious facts of foresight and purpose. Their evolution is not entirely or even mainly naturalistic: it involves a spiritual factor which is manifested clearly in the history of man, and which is as distinct from natural selection as natural selection is distinct from the mechanical causes to which inorganic changes are referred. The method of evolution begins with mechanism, is changed when life appears, and changed again when life becomes self-conscious and can look before and after. At each of these stages the appearance of a new factor can be observed, and this new factor affects the result and modifies the method of the process. It is characteristic of naturalism to attempt an interpretation of the whole process in terms of those factors only which appear at the earlier stages. The theory has the advantage of simplicity; but the simplicity is gained at the expense of the facts."

It is when this purposive direction of moral evolution begins that man has to ask not merely how moral ideas and institutions have been arrived at, but what is the line along which progress is to be desired; he has to ask what course *ought* to be taken, in other words, what is good?

"The further we go in examining any naturalistic theory of ethics," says Professor Sorley,[1] "the clearer does it become that it can make no nearer approach to a solution of the ethical question than to point out what courses of action are likely to be the pleasantest, or what tendencies to action the strongest; and this it can only do within very narrow limits both of time and accuracy. As to what things are good it can say nothing without a previous assumption identifying good with some such notion as pleasant or powerful. The doctrine of evolution itself, which has given new vogue to Naturalism, both in morality and

---

[1] *Ethics of Naturalism*, pp. 326–327.

in philosophy generally, only widens our view of the old landscape. By its aid we cannot pass from 'is' to 'ought,' or from efficient to final cause, any more than we can get beyond the realm of space by means of the microscope or the telescope."

The manner in which a system of ethics of evolution becomes possible when reality is explained as "depending upon and expressing mind," is so lucidly stated in the concluding paragraphs of his work that I shall give it in his own words:—[1]

"The character of the course of evolution is seen in a different light when it is recognised that human conduct and its methods must be taken into account in interpreting the process. The scientific writers who have been most forward in pressing the claim that man must be held to be a part of the cosmic process have also, unfortunately, been inclined to interpret the whole process, not as it is, but as it would be apart from human intervention and the ideals which man brings to bear upon it. But the claim that man must be interpreted as part of the universe involves the counterclaim that the nature of the universe cannot be understood apart from the distinctive features of man's activity. And, when this is allowed, the naturalistic interpretation of evolution becomes increasingly difficult. Evolution can no longer be regarded as entirely purposeless, for that part of it which we call human conduct undoubtedly displays purpose. It cannot be entirely indifferent or antagonistic to morality, for the action of men, which enters into the process, bears the impress of moral ideas. These considerations are not put forward as proving the truth of the view that the process of evolution is the expression of a divine purpose. They prove only that purpose and intelligence are somewhere within the process, not that they are present everywhere, or that the whole course of nature is the expression of one increasing purpose. But the facts leave room for this interpretation, if they do not demonstrate it. Its justification would require the establishment of a view of the world which may be called idealist, seeing that it would explain reality as depending upon and expressing mind. No such justification can be attempted here. But it may not be out of place to remark that it enables us to avoid both the fruitless efforts of the naturalists to derive an ethical doctrine from the history of development, and the antagonistic view urged by Huxley that the cosmic and moral orders are in hopeless conflict. It avoids the latter view, because it regards the moral ideas and institutions of man as part of the complete process, as factors in the movement which leads in time from nature to spirit. And it avoids the former view because it holds that the ethical element which is manifested latest in the temporal process is presupposed from the first and necessary to the understanding of the whole. The ideal of goodness many contribute towards the interpretation of evolution, but its own explanation must be sought by another method."

We see then that the theory of primal intelligence propounded in this chapter affords an adequate solution of every one of the

[1] *The Ethics of Naturalism*, pp. 331–333.

problems which have been found incapable of expression in terms of the mechanical and molecular schemes, with the single exception of the origin of energy. Moreover, this has been effected without attributing to the primal intelligence any powers other than may be derived by the legitimate physical process of inference from observed data. It still remains for us to inquire whether any alternative scheme is offered which purports to account for the phenomena. The purely materialistic position which denies the existence of mind, except as a function of matter, is now so completely recognised as "not so much an error as an absurdity," to quote Sir Leslie Stephen's words, that no thinker can be found to champion it, and the latest scheme offered as an alternative to primal intelligence is, as far as I am aware, one propounded by Professor Haeckel under the name of *monism*, a term originally employed by Wolff to denote either the pure materialism which denies any reality to mind, or the pure idealism which denies any reality to matter. Professor Haeckel is a zoologist who has acquired a deservedly high reputation by his important contributions to the knowledge of the more lowly living organisms, as embodied in his monographs on the *Radiolaria*, *Sponges*, and *Jelly-fishes*. He has also become known to a large circle of readers by his popular works on biological evolution. These are written in an interesting style, and some of them have been translated into English, but from the point of view of the uninstructed who read them for information, they all appear to be subject to the defect of not always distinguishing between statements of ascertained truth and mere unsupported hypotheses. Even such hypotheses, or guesses at truth, may be of considerable interest when dealing with subjects on which the writer can claim to be an authority, but it is confusing to the unlearned when they are put forward, without any reservation, as facts. A noteworthy example of this is to be found in his genealogy of man,[1] in which, not being content with the dimly discernible main outlines indicated by observation, he fills in the details with heraldic minuteness, even going to the length of inserting imaginary types of organisms which he thinks likely to have existed, but which no observer has ever seen either in living or fossil form. Moreover, these types are presented as though they represented ascertained facts, without any indication of their purely hypothetical character. It is not, therefore, surprising that the great physiologist and secretary of the Berlin Academy of Science, du Bois Reymond, should have described this genealogy as having about as much value as the genealogies of Homer's heroes to a critical historian. We

[1] *History of Creation*, vol. ii., chapter xxvii.

shall see that when Professor Haeckel comes to deal with physical phenomena, a subject in which he is evidently not at home, this fault becomes much more pronounced, and not only unsupported hypotheses, but hypotheses which, to any well-informed student of physics, are obviously absurd, are put forward as facts.

The fundamental assumption of Haeckel's scheme is the absolute purposelessness of all natural activity, which he sets forth as follows:—[1]

"*Dysteleology, or the theory of purposelessness*, is the name I have given to the science of rudimentary organs, of suppressed and degenerated, aimless and inactive, parts of the body; one of the most important and most interesting branches of comparative anatomy which, when rightly estimated, is alone sufficient to refute the fundamental error of the teleological and dualistic conception of Nature, and to serve as the foundation of the mechanical and monistic conception of the universe."

As I cannot claim any authority as a critic of biological hypotheses, I append the following extract from Professor Huxley's review of the original German edition:—[2]

"Professor Haeckel has invented a new and convenient name, 'Dysteleology,' for the study of the 'purposelessnesses' which are observable in living organisms—such as the multitudinous cases of rudimentary and apparently useless structures. I confess, however, that it has often appeared to me that the facts of Dysteleology cut both ways. If we are to assume, as evolutionists in general do, that useless organs atrophy, such cases as the existence of lateral rudiments of toes in the foot of a horse place us in a dilemma. For, either these rudiments are of no use to the animal, in which case, considering that the horse has existed in its present form since the Pliocene epoch, they surely ought to have disappeared, or they are of some use to the animal, in which case they are of no use as arguments against Teleology. A similar but stronger argument may be based upon the existence of teats, and even functional mammary glands, in male animals. . . . There can be little doubt that the mammary gland was as apparently useless in the remotest male mammalian ancestor of man as in living men, and yet it has not disappeared. Is it, then, still profitable to the male organism to retain it? Possibly, but in that case its dysteleological value is gone." In later editions of the works containing the reprint, Huxley makes the additional observation that "the recent discovery of the important part played by the thyroid gland should be a warning to all speculators about useless organs."

The further development of Professor Haeckel's scheme, if it can be dignified by such a name, has for its basis a long series of hypotheses propounded in the form of positive assertions, but

[1] *History of Creation*, vol. ii. p. 353.

[2] *The Academy*, 1869; reprinted in *Critiques and Addresses*, and in *Darwinismus*.

without any attempt to establish their validity. One of these is the principle of the conservation of energy, which is of course admissible; another is the indestructibility of matter, which, although now known to be untrue in the sense formerly ascribed to it, becomes admissible when the term matter is extended, as it must be, to include its ethereal basis. Then follow two unsupported hypotheses, the first of which is that the universe is infinite in its spatial extension, which considerations dealt with in Chapter XXIII. show to be improbable, although its falsity cannot be said to be definitely demonstrated. This question, however, is not material to his argument. The second is that the universe is eternal, having neither beginning nor end, and this is in direct conflict with the observed dissipation of energy, and, with respect to the purposeless universe assumed by Haeckel, is demonstrably untrue, in consequence of the observed molecular constitution of matter. Of two further assertions which attain the acme of physical absurdity, one denies the law of the dissipation of energy, the only reason advanced by Professor Haeckel being that its validity is not consistent with his hypotheses; the other sets aside the law of inertia of matter, and replaces it by the assertion that ether and matter possess sensation and will, likes and dislikes, and the capacity of directing their action in accordance therewith. Finally, Professor Haeckel assumes that the continuous strife, by every atom in the universe, to follow the course mapped out for it by its own desires, affords an adequate foundation for the observed order of nature. It is obvious that a scheme which is founded in this manner upon unverified hypotheses, and which not only makes no attempt to verify them, but, when they are in conflict with observation, rejects the evidence in order to retain the hypotheses, is altogether unscientific, and does not therefore call for detailed consideration in a physical treatise. That such a production should emanate from a mind which has been the source of noteworthy contributions to scientific knowledge presents an interesting, but by no means unique, problem in mental pathology.

Were it not that the *Riddle of the Universe*, which is devoted to the exposition of this scheme, has been largely circulated, and regarded as a scientific authority, amongst readers whose training is insufficient to enable them to distinguish between argument and mere assertion, it would have been passed over in silence. To enable readers to satisfy themselves that the foregoing presentation is not a mere travesty, Professor Haeckel's fundamental hypotheses are enumerated in his own words in Appendix Q. Moreover, the English translation of the work itself is obtainable in a sixpenny edition.

It may be of interest to the reader, before bringing this chapter to a close, to indicate very briefly the line of thought which led the writer to the formulation of the theory of the universal mind which he has endeavoured, in the preceding pages, to establish on a strictly scientific observational basis. His attention was first directed to the relations between mind and energy when considering the strikingly and unexpectedly successful result of a very simple experiment made to test the reality of the alleged possibility of thought transference between mind and mind, other than through the recognised sense channels. This result did not indicate any more direct action than in the ordinary methods where these channels are utilised, and was therefore of no assistance in the solution of the problem with which the present chapter is concerned. It showed, however, quite conclusively, the existence of physical channels of communication between brain and brain, other than the senses generally recognised, and, on account of its intrinsic interest, is described, and its bearings are discussed, in Appendix R. Attention once directed to the relations between mind and mind, it was only natural to consider them in connection with that unity of nature which is indicated by the whole trend of modern scientific progress. In the days of the old Greek philosophers, or seekers after wisdom, the sum-total of nature knowledge was so restricted that the whole realm of natural phenomena was the domain of each and every one of them, and this state of things continued on into the Middle Ages. It was not until the sixteenth century was well advanced that the philosopher began to specialise, and this specialisation attained its maximum development during the eighteenth, and the earlier portions of the nineteenth century, when scientific knowledge appeared to be divided into a cluster of walled-in compartments. Then began the era of generalisation, in the development of which the two great physicists Hermann von Helmholtz and Lord Kelvin were pre-eminent above all others, while from Darwin and Wallace came the important contribution which made it possible to construct a scientific scheme of living structure, and one, moreover, in closest harmony with the unity indicated by the physics of the inorganic. The conviction of this ultimate unity of all physical phenomena is driven home with overwhelming force by the modern theory of electricity and matter, according to which the sum-total of the varied material phenomena, exhibited by living as well as non-living matter, forms a single all-embracing unity consisting ultimately only of ether in motion, or, in more abstract language, of energy changes. If, then, the minds of men and other organisms were discrete entities, without any correlation, they would constitute, as far as we can ascertain, the only

instances of discreteness in the physical universe, and their correlation through some such unity as that of the universal mind is philosophically indicated by the principle of continuity. Moreover, such discrete units unconnected with the All of material phenomena could play no active part in the latter, and we should be thrown back upon the imperfect representations of the phenomena of consciousness which have been discarded as repugnant to our reason. Finally, we should find ourselves in face of the insoluble problem of accounting for representations of material phenomena, attained only through these discrete entities, unconnected with the All, and yet blending into this harmonious unity, arrived at by the operations of many minds.

From this point of view, the existence of a universal mind, correlating the otherwise separate units forming the minds of all living organisms, appears as an inevitable conclusion, and the only question to be determined is whether it is to be regarded as the origin, or as the result, of evolution. The former alternative accounts for the formation and maintenance of the universe, in accordance with the observed laws of nature, on the simple assumption of an enlargement of the observed content of the universal mind, without any change in its nature. The latter alternative does not. In other words, the theory of primal intelligence is verified by the ultimate scientific criterion of its adequacy to account for the observed facts, while the hypothesis of the development of mind in the course of evolution is found to be inadequate. The scientific thinker, therefore, finds himself compelled to reject the latter and to accept the former. The necessary correlation of the minds of all living organisms then finds a simple and natural explanation in the fact of their common derivation from, and continuous relation to, the universal mind or omnipresent intelligence.

It is necessary that the limitations of the argument here developed should be recognised. In the first place, it does not, at the stage here attained, provide a proof of the creation of the universe, but only of its formatiom, or building up, from the material available, the total energy-content. It was found necessary to assume the existence of this content in order to account for the material universe as presented to our experience, although the nature of the initial distribution of the energy was seen to be immaterial. There does not appear to be any inherent philosophical impossibility of such further extension of our empirical data as would enable us to deduce the creation of the energy itself by the power of mind. If we could make this further step, the purely physical argument would lead immedi-

ately to the definite demonstration of the existence of a creative God, as presented in the many forms of popular theology. The relation definitely established between mind and energy, however, in that it completely subordinates energy to mind, strongly suggests this conclusion. Moreover, if mind is not itself the higher unity within which energy is embraced, and from which it proceeds, the argument of the Hegelian Logic, as I shall presently show, affords absolute demonstration, on the basis of the relation established between mind and energy, that if the latter is not included in the former, both must necessarily be included in a higher unity. The philosophical principle of economy will prevent us from seeking an explanation of the universe in terms of such a higher unity unless it be found that mind is insufficient, and any such higher unity must necessarily be something greater than, and including, mind in all its fulness. In either case we see that the physical argument leads to a complete resolution of the Cartesian and Kantian dualism between mind and matter, between the percipient subject and the perceived object. It therefore effects a complete rationalisation of the whole scheme of the universe as known to us. Reason, the highest principle of mind as we know it, is shown to reign supreme throughout the universe, subject only to the possible existence of higher principles representing aspects of the ultimate reality which are beyond the reach either of our powers of observation or of our reason.

The very limitations inherent in the nature of this argument constitute much of its value. In place of a direct presentation of the self-existent, self-determined, and therefore infinite, ultimate reality, which our reason is, in consequence of its own limitations, totally incapable of comprehending, it leads to a presentation limited indeed, but only to the extent of our own powers of comprehension. We have presented to our cognition only those aspects of the ultimate reality which are in direct relation with ourselves, and only to the extent to which our limited reason is capable of cognising them. Moreover, the knowledge of the nature of causality which physical science in its present stage of development places at our disposal makes it possible for reason to take the one further step open to it, and to infer with certainty the self-existence and self-determination, and consequently the absolute infinity, of the ultimate reality of which we have already learned that reason, if not the highest aspect, is the highest which can possibly become an object of cognition.

We have seen that the isolation of a phenomenon, or even of a group of phenomena, can never be complete, so that when we select two phenomena, or groups of phenomena, and regard one as

the cause of the other, such a representation is necessarily imperfect, and can, under the most favourable circumstances, only lead to a first approximation towards a solution of the problem of determining their mutual relations. The complete cause of even the simplest phenomenon is nothing less than the All. Such partial representations of the principle of causality cannot, therefore, afford an adequate basis for the wide generalisations of metaphysical inquiry, and their employment for such purposes is necessarily illegitimate, and may be expected to lead to apparently irresolvable contradictions, or *antinomies*.[1]

A very great part of the immense service rendered by Kant's critical philosophy, as formulated mainly in the *Critique of Pure Reason*, lay in the incisive manner in which the *Antinomies of Pure Reason* were there displayed in the strongest relief. Kant never carried out his argument to its ultimate logical conclusion, and it is difficult to avoid the suspicion that he must have been deterred from doing this by the impossibility of that conclusion, which is nothing less than absolute philosophical nihilism, consisting in the negation of the possibility of knowledge of any kind.

It was definitely shown by Kant that existence can only mean existence for consciousness, existence unrelated to a conscious self being a mere empty abstraction. He failed, however, to perceive the necessary relation between phenomena and noumena, between physical and metaphysical reality. He endeavoured to escape from the region of *Speculative Reason*, in which all appeared to be determined by external laws, the observed laws of nature to which all phenomenal relations are found to be subject, into the region of *Practical Reason*, the region of freedom, or self-determination. To effect this he utilised the principle that the intelligence consists, not merely in consciousness, but in self-consciousness. In virtue of the former it is externally determined, and so apprehends itself as belonging to the world of phenomena; while in virtue of the latter it is self-determined, and so possesses the consciousness of belonging to a world of freedom, the world of noumena. Kant finds the necessary basis for establishing the reality of the region of the *Practical Reason* in the categorical imperative of conscience. The law of the moral consciousness would appear to be in direct contradiction with the self's empirical consciousness of itself, but it could not on that account be regarded as an illusion, since Kant had previously shown that the self's knowledge

[1] An *antinomy* may be defined as a pair of mutually contradictory propositions, each of which, when considered apart from the other, appears to be demonstrably true.

of itself as an object of experience cannot be legitimately applied to the knowing subject. The objection on the ground of phenomenal determinism is therefore removed by the proof that the knowing self is not a phenomenon but a noumenon. The supremacy of the practical reason is thus established, and a place found for the freedom of the intelligence without questioning the truth of phenomenal determinism.

Such a correlation of the regions of the speculative and practical reason is, however, purely superficial, and establishes only an external harmony between them, and, indeed, practically implies the impossibility of their true unification. There remains, therefore, an irreconcilable dualism between the intelligence and the phenomenal world. This was made far more definitely evident by the subsequent development of the two aspects of Kant's system by Fichte and Schelling. Fichte, in his endeavour to exalt the intelligence at the expense of the world of phenomena, reduced the latter to a practical nonentity as a mere object of thought created by the self. Schelling, on the other hand, was ultimately led to reject idealism completely, and to seek the ultimate reality in a unification of mind and matter which was attained by practically ignoring their differences; and this ignoring of the most fundamental of distinctions necessarily involved the ignoring of all distinctions. The ultimate reality was therefore reduced, according to Schelling's own statement, to a "pure indifference."

This complete demonstration of the irreconcilable nature of the Kantian dualism led Hegel to attempt the real unification of Nature and Intelligence by proving that their very antagonism is itself a manifestation of their unity, and so to conserve the freedom belonging to man as a rational and moral being, not by removing it to a region beyond the grip of physical necessity, but by showing that it realises itself in and through that necessity itself. The divergent developments of the Kantian idealism by Fichte and Schelling had shown that it could only be maintained by proving mind to be the key to the explanation of material phenomena. An analysis of mind, even more fundamental than was achieved by Kant, was therefore the first essential to the solution of the problem. The nature of the key to this first stage of the solution, though discovered by Hegel from a purely metaphysical standpoint, will be more easily and completely apprehended by the physical student if it be considered from the physical standpoint which the development of science has now rendered possible.

This standpoint is simply the concept of the nature of causality in nature, that is to say, in the world of phenomena, to which

science has conducted us, and which I have endeavoured to make plain throughout the present volume. We have learned that the old concept of a chain of causality, correlating a group of otherwise discrete phenomena into a linear series, is a representation too incomplete to be of service beyond the first, but necessary, stage of a scientific investigation, in which we are seeking for a first approximation to the actual relation.

Hegel's great achievement was to show that the causal relation existing between the categories of the understanding is of exactly the same character as the law of causality now known to hold good in the material universe of physical science. The procedure employed is analogous in every respect to that which we have found requisite in the investigations of physical science. The several categories are first considered, for the purpose of a first approximation, as separate entities. It is then shown that, although this stage cannot be dispensed with, it leads inevitably to contradictory results, which can only be eliminated, or rather resolved into higher verities, when it is realised that the different categories are not isolated ideas and mere instruments of thought, but are elements in a complex whole, or stages in a complex process, and that the mind in its unity is itself this complex whole or complex process.

The Hegelian Logic is developed in three definitely marked stages. The first, starting with the category of Being, is confined to the consideration of the categories which, though implicitly involving relativity, do not express it explicitly. Hegel compares this first attitude of thought with the pre-scientific consciousness of natural phenomena. The second stage, corresponding to the scientific, or reflective, consciousness, is concerned with those categories in which relativity is explicitly expressed. In the third stage, which corresponds to the philosophical consciousness, Hegel adds a new genus to the genera included in Kant's enumeration of the categories, embracing the categories of *Ideal unity*, and it is then shown that all the categories form a system, or complex, of essentially related transitory existences, each of which exists only as determining, and being determined by, the remainder, in accordance with universal laws. The categories are, therefore, ultimately left to define themselves, simply by the necessary movement of thought through which they carry the mind, and it is then found that they lead to the idea of self-consciousness as their ultimate meaning or truth. The intelligence then realises itself in the phenomenal world by discovering that it lies in a world of objects, each and all of which exist only in so far as they exist for intelligence, and in so far as intelligence is revealed or realised in them.

Hegel then attempts, in the *Philosophy of Nature*, to prove that the laws of the phenomenal world are, in their essential character, identical with the laws of the intellect, in which case Nature as well as Intelligence would find its ultimate expression in terms of self-consciousness. In this attempt, it must, I think, be admitted, he was only partially successful, the development of physical science not having at that time reached a sufficiently advanced stage to make it possible. Unless, however, the validity of the argument developed in this chapter can be impugned, the stage of development now attained by physical science has made it possible to carry the Hegelian argument to the final stage of complete demonstration.

From the point of view here attained the whole physical universe of mind and matter may be compared to a living organism the central principle of which is a self-determining Being. The independence of the constituents of this organism must not, however, be such as to permit of any break in their connection with each other, which would destroy the unity of the whole. Such a universe will appear under one aspect as a unity in which there is no place left for difference or distinction; under another aspect it will, however, appear to be separable into an indefinite number of fragments, each seeming to be centred in itself, or self-determined. This is Hegel's conception of the universe, and in reference thereto, Professor E. Caird observes:—[1]

"Under such a conception the usual antithesis of individualism and pantheism fails us, and our idea of the world seems to involve both at once, or to fall into a kind of alternation between them, such as is found in the monadism of Leibnitz, or in the later theory of Schelling, in which all the differences of things were said to be 'not qualitative, but merely quantitative,' *i.e.* to be differences that from the highest point of view might be neglected as unessential. This, however, were to forget that though the organism is organic in all its parts, yet these parts have their specific determination, and that it is through this specific determination that they form one whole. It were to forget that though a self-determining principle necessarily is present in its determinations, and gives them thus a certain independence, yet that they in turn are limited in themselves, and only maintain themselves as the principle realises itself in them, or, in other words, as they in turn surrender themselves to the life of the whole. Their capacity of so surrendering themselves, in short, is the measure of their reality. Thus the unity as a self-determining principle is *in* the differences, but it is also in their negation, by which they pass beyond themselves as individuals and so return into unity."

We see that the concept of the ultimate reality here developed is not that of a unity from which all distinctions, and therefore

[1] *Hegel*, by E. Caird, p. 180.

all relations, are eliminated. A unity of that kind is now seen to be a mere empty abstraction, so that there is no meaning either in its affirmation or denial. Kant's demonstration, in the *Critique of Pure Reason*, that such a unity can neither be affirmed nor denied is irrefutable, and we can now see that any other conclusion would be absurd. Kant's error lay in the assumption that ultimate reality was to be found in pure self-identity entirely divested of differences. The same concept is reproduced in the "unknowable" of Herbert Spencer's philosophy. The ultimate reality of Hegel, which is now seen to be identical with that indicated by modern science, consists in a self-determining unity which reconciles and includes within itself all the diversities and apparent contradictions of both Intelligence and Nature. In place of relativity being excluded from the Absolute, *all* relativity is included within it, but the sphere of its relations is entirely contained within itself and determined by itself.

It will be of interest to conclude with a brief reference to the most ancient of all known systems of philosophy, the philosophy of the Vedas, which forms the essential basis of Brahmanism and of its great offshoot Buddhism, and to note how its fundamentally different method of dealing with the same great problem affords further confirmation of the validity of the conclusion to which we have been led. Starting from the phenomenal system of nature, we have worked our way upwards to the conception of the Absolute as self-consciousness of a completeness which we can only apprehend in partial limited aspects. The Eastern student of philosophy subjects himself to a long and severe course of bodily and mental training designed to enable the higher elements of the intelligence to obtain control both of the bodily functions and of the lower elements of the intelligence itself. It is with this sole object that the Brahman, after having fulfilled the ordinary duties of a citizen and of family life, retires to a life of abstinence and contemplation, seeking to realise the self, that is to say, to attain to that higher stage of self-consciousness in which subject and object are both actually realised as essentially one, not only with each other, but with the Eternal Self-consciousness constituting the ultimate reality known under the designation of Brahm. When this is once realised, the Absolute is recognised as being the essential reality, in comparison with which the lower reality of the phenomenal world is but an illusion. This realisation necessarily involves the recognition of his own true self, freed in mental vision from its phenomenal environment, as identified with and included within the higher self-conscious unity, or ultimate reality, of which the universe of

nature is but the phenomenal manifestation. Their starting-point is our goal.

The theory of the universe, as indicated in the Vedas and Upanishads, is developed by purely *a priori* reasoning from this conception of the Absolute. The Eternal Self-consciousness, in the perfect freedom of its absolute self-determination, first manifested itself in the form of mind. From this first manifestation as mind there was evolved a second manifestation in the form of force. Force in its turn gave rise to ether in which vortices, or "whirlpools," as the Vedas express it, were formed, giving rise to a series of ethers of increasing density and complexity. Matter as we know it was supposed to be derived, by further increase in complexity, from the final term of the series of ethers.

There can be little doubt that the Greek speculations referred to in Chapter IV. derived their origin from this source, and that the complex system of ethers of pre-Newtonian speculative physics, derived directly from the Greek philosophers, was therefore of Aryan origin. These speculative ethers formed a series of sinks for the disposal of every unexplained physical difficulty, and were only disposed of by the investigations of Galileo, Huygens, Newton, and their successors. There can be no doubt that the acceptance of the wave theory of light was retarded for at least a century by the fears of the lesser physicists lest the scientifically developed ether theory of Huygens should lead to the recrudescence of the miasmatic fog of complex ethers which, they hoped, had finally disappeared from serious physics. The ether theory adopted by Professor Haeckel in the *Riddle of the Universe*, which he attributes to a modern writer named Vogt, is nothing else than a return to these discarded speculations, and is not susceptible of any relation with serious scientific theory. It affords a striking example of inconsistency that Professor Haeckel should attempt to construct a theory of the universe from which the primal intelligence is to be excluded, not on the basis of the ether as known to science, that is to say, the ether of which the properties are derived from empirical observation; but on the purely *a priori* basis of a pre-Newtonian ether, with properties inferred by introspective contemplation from the concept of the Eternal Self-consciousness as the ultimate reality.

Although the Vedic scheme of the universe does not, and could not be expected to, bear detailed comparison with that of modern physical science, the similarity of their broad outlines is extremely remarkable, even at first sight. It becomes still more so when we take note of the fact that the very indefinite Vedic concept designated by the term force corresponds far more with the

modern concept of energy than with that of force; and when the term energy is substituted for force, the Vedic scheme of development becomes identical with the one which expresses the most recent developments of physical research, viz. :—

The Absolute, or Eternal Self-consciousness.

Mind.

Energy.

Ether.

Matter.

Vedic Scheme. ↓

Scientific Scheme. ↑

MATHEMATICAL AND OTHER APPENDICES.

APPENDIX A.

PRINCIPLES OF THE VARIATIONAL METHOD EMPLOYED FOR DERIVING THE EQUATIONS OF EQUILIBRIUM, OR OF MOTION, OF A DYNAMICAL SYSTEM.

THE analytical method known as the *Calculus of Variations*, although it has been considerably amplified by later mathematicians, has undergone no alteration in essential principles since it passed through the master mind of Lagrange, who so vastly developed and amplified it from its previously existing germs, that he is justly and properly regarded as its inventor. Like most of the more important advances in mathematical analysis, it arose, however, from the generalisation of mathematical methods devised with the object of obtaining the solution of a problem for the direct attack of which previously existing methods were inadequate.

It was in the month of June 1696 that John Bernoulli propounded the celebrated *Brachistochrone Problem*

"to determine the nature of the curve joining two points in the same vertical plane, such that the time of descent from the higher to the lower one, of a particle moving under the action of gravity, should be less than for any other possible path between them."

The propounder himself gave what may be described as an accidental solution, derived from an analogy based on what is known as *Fermat's law of least time*.

The Greek mathematician Hero had deduced the ordinary law of reflection at a plane surface from the hypothesis that the path was the shortest possible one connecting the object and its image through a point on the reflecting surface. Fermat, early in the seventeenth century, sought to deduce the law of refraction in a similar manner, but found that the path followed by a refracted ray passing from one medium to another of different density was

not the shortest route, but the one which occupied the shortest time. Huygens subsequently showed that Fermat's law might be extended to the passage of light through media in which the density varied continuously from point to point. By means of this analogy, John Bernoulli obtained a beautifully simple solution of his Brachistochrone Problem. His brother James gave a much less elegant solution of this problem, but developed for the purpose a general geometrical method applicable to a whole class of analogous problems included in the following general enunciation of what is known as the *Isoperimetrical Problem* :—

"Of all isoperimetrical curves, that is to say, curves of equal perimeters or lengths, between the same two fixed points, to find the curve such that the space included (1) by a second curve, each of whose ordinates is a given function of the corresponding ordinate, or the corresponding arc, of the curve sought, (2) by the ordinates of the extreme points, and (3) by the part of the axis of abscissæ lying between those ordinates, shall be a maximum or a minimum."

James Bernoulli's method involved the assumption that :—If a curve has a certain property of maximum or minimum, then every portion of the curve has the same property. Euler observed that, though this assumption was admissible in the case of the brachistochrone in a vacuum, it was not so when the vacuum was replaced by a resisting medium. The entire length and form of the preceding path is then involved in the determination of the velocity in any element, so that the whole curve can be brachistochronous without the separate elements necessarily exhibiting the same property. This observation led Euler to very considerable generalisations, based on substantially similar principles to those afterwards employed in the methods developed by Lagrange, but Euler's methods remained essentially geometrical, which necessarily limited their scope, and also made them exceedingly difficult of application. Euler, moreover, made no attempt to extend the method to the treatment of surfaces having maximum or minimum properties, that is to say, to cases in which the property in question is expressed by a double integral.

The great advance of providing an analytical foundation for what is essentially an analytical method, and thereby freeing it from the limitations of a geometrical framework, was made by Lagrange. This great mathematician noticed that the increments of a function, due to changes in the form of the function, are of similar character to those due to changes in the independent variables, and for the former he employed the symbol δ, reserving the symbol d for the latter. This principle, which forms the foundation of the Calculus of Variations, proved to be an extremely fruitful one in the hands of Lagrange, but he never attempted to

give a demonstration of its validity, which he seems to have left to be inferred from the successful results of its application to the solution of a wide range of problems. This procedure resulted in his contemporaries and successors finding the greatest difficulty in clearly grasping the distinction between a variation and a differential, and there can be no doubt that this difficulty was considerably enhanced by the fact that Lagrange did not always in his own work show a clear logical discrimination in his use of the two symbols.

Euler endeavoured, unsuccessfully, to provide a means of distinguishing between a variation and a differential, by assuming the change of form of a function to arise from changes in constants contained in it, and defining the variations as the increments due to these changes. This conception is a very confusing one, and greatly restricts the scope of the calculus. It is pointed out here because its influence is shown in at least one comparatively recent work. The difficulty was first cleared up, as far as I am aware, in Jellett's excellent treatise,[1] which has now been long out of print.

Quantities generally may be divided into constants and variables, and the latter again into independent and dependent variables. Independent variables are those which may have arbitrary values assigned to them, while dependent variables are such that their changes depend on those of independent variables connected with them by some kind of relation. The dependent variables are said to be functions of the independent ones, and the nature of the relation is termed the form of the function. Such a relation is usually expressed by the notation

$$u = \phi\,(x_1\,,\,x_2\,,\,x_3\,,\;.\;.\;.\;.\,)\,,$$

where x_1, x_2, x_3, etc., are the independent variables, u is the dependent variable, and ϕ is a general symbol denoting the form of the function.

Jellett points out that a change in the value of the function may be effected, either by changes in the values of the independent variables, or by a change in the form of the function, from which it follows that functions may be divided into two classes, corresponding to constant and variable quantities, respectively; viz. functions of constant, and of variable, form, respectively. Jellett calls the former *determinate*, and the latter, *indeterminate* functions. I prefer to call them *determined*, and *undetermined*, functions, respectively, for it appears to me logically objectionable

[1] *An Elementary Treatise on the Calculus of Variations*, by the Rev. J. H. Jellett, Dublin, 1850.

to describe as indeterminate a function which can become determined, partially or completely, by the introduction, with the aid of a suitable mathematical procedure, of the conditions proper to the problem which may be under consideration. Determined functions can change their values only through changes in the values of one or more of the independent variables, and the rules for determining these changes form the subject of the Differential Calculus. Undetermined functions are susceptible of similar changes, and are, in addition, susceptible to changes arising from changes in form. It is this latter class of changes which is peculiar to the Calculus of Variations.

The form of a function may evidently be related in such a manner with the forms of one or more other functions that its form may be expressed in terms of theirs. For example, the form of the differential coefficient of a differentiable function may be expressed in terms of the form of the function itself. In such cases Jellett gives the name of *primitive functions* to those whose forms may be arbitrarily assigned, while one whose form depends on those of the primitives he calls a *derived function.* He denotes a derived function of a primitive ϕ by the symbol $F . \phi$, and it must be noted that $F . \phi$ is not a function of ϕ, the symbol F denoting a relation between forms, not between magnitudes. The general problem of the Calculus of Variations may then be defined as the determination of the change in value of $F . \phi$, resulting from a given change in the form of ϕ; that is to say, the determination of the change of value of a determined function F of an undetermined function ϕ, due to a change in the form of the undetermined function.

Just as in the problems dealt with by the Differential Calculus, it is essential that the increments assigned to the variable quantities should be capable of indefinite decrease for all values of the independent variables, so that if ϕ changes its form to ϕ_1 and i is a constant quantity of whatever degree of smallness $\phi_1 - \phi$ is to have, the quantity $(\phi_1 - \phi)/i$ must be finite for all values of the variable consistent with the conditions of the problem. We may therefore write $\phi_1 - \phi = i\psi$, where ψ is a function of the same variables x_1, x_2, x_3, etc., which must not become infinite for any admissible values of the variables, and the same restriction must apply to any derived functions $F. \phi$ which enter into consideration. The increment $i\psi$ is said to be a *variation* of the function ϕ, and a variation may therefore be defined as an indefinitely small change in value of a function which arises from a change in its form.

It is clear from the preceding that the variation of a primitive function is entirely arbitrary, and that the variation of a derived

function depends on that of its primitive. If $u = \phi(x_1, x_2, x_3, \ldots)$ be an undetermined function, and $V = F.u$, a function derived from u, then the variation of V may be obtained by substituting $\phi + i\psi$ for ϕ, and performing the operation F on the function so obtained, so far as is required to obtain the co-efficient ω of the first power of i. The required variation will then be $i\omega$. This follows by reasoning precisely similar to that employed in the Differential Calculus for the derivation of the increment of a function.

This statement of the problem is more general than is requisite, since in practice we are concerned only with functions derived from their primitives either by successive differentiation or integration. Now, all functions so derived are easily seen to satisfy the distributive law for the operation F, so that we always have

$$F.\phi + F.\phi_1 = F.(\phi + \phi_1).$$

We may, therefore, simplify the method given above for finding the variation, viz. in $F.\phi$ substitute, as before, $\phi + i\psi$ for ϕ. Then

$$F.(\phi + i\psi) = F.\phi + F.i\psi,$$

so that the total variational increment of $F.\phi$ is $F.i\psi$. But it is easily shown to be a consequence of the distributive law that $F.i\psi = iF.\psi$, which therefore represents the total variational increment of $F.\phi$.

In accordance with Lagrange's notation, let du denote the differential increment of a function u of any number of variables, being the increment due to changes in the independent variables, δu the variational increment, due to changes in the form of the function, and Du the total increment. It then follows, by reasoning similar to that employed to demonstrate the validity of the principle of the superposition of small motions, that

$$Du = du + \delta u.$$

If u is a determined function of variable quantities, $\delta u = 0$, so that

$$Du = du;$$

while if u is an indetermined function of constant quantities, $du = 0$, so that

$$Du = \delta u.$$

An independent variable can receive only one kind of increment, and therefore it is immaterial, as far as the attainment of correct

conclusions is concerned, what symbol is employed to express it. Lagrange appears to have used the symbols d and δ indifferently for this purpose, and there can be no doubt that as the distinction between differentiation and variation had not then been definitely expressed, this increased the confusion in the minds of many of his contemporaries and successors as to the meaning of his method. Since such an increment is denoted by d in the Differential Calculus, and is not altered in meaning when denoted by δ, it is in every way preferable to adhere to the former symbol.

Moreover, we shall see presently that, in the mechanical applications of the Calculus of Variations, the symbol δ is applied to the independent variables constituting the co-ordinates of a particle, being then employed to denote a displacement of a particle in space, or, as it is sometimes called, a mechanical variation. The increments due to small displacements can, in general, be determined by the procedure of the Calculus of Variations, for reasons which will be pointed out; but obscurity and confusion will be introduced into the subject unless this special employment of the symbol is carefully distinguished from its use in purely geometrical problems. When employed to represent a mechanical variation, the symbol δx has a meaning essentially distinct from that represented by dx, and this makes it the more important that the symbol δx should not be employed to represent the concept usually symbolised by dx, even in cases where it could obviously have no other signification.

(1) Let u be a primitive undetermined function

$$u = \phi(x_1, x_2, \ldots . x_n).$$

Then

$$\mathrm{D}u = \left\{ \begin{matrix} du \\ + \\ \delta u \end{matrix} \right. = \left\{ \begin{matrix} \dfrac{\partial u}{\partial x_1} dx_1 + \dfrac{\partial u}{\partial x_2} dx_2 + \cdots + \dfrac{\partial u}{\partial x_n} dx_n \\ + \\ i\psi(x_1, x_2 \ldots . x_n), \end{matrix} \right.$$

the symbols d and ∂ having the usual signification of complete and partial differentiation, respectively.

(2) Let

$$u = \mathrm{F} . \phi(x_1, x_2, \ldots x_n),$$

where F obeys the distributive law.

Then du has the same value as in (1), and

$$\delta u = \mathrm{F} . i\psi = \mathrm{F} . \delta\phi,$$

(3) Let

$$\mathrm{V} = f(x_1, x_2, \ldots . u_1, u_2, \ldots .),$$

a determined function of x_1, x_2, etc., and u_1, u_2, etc., of which the former are independent variables, and the latter are undetermined functions of any or all of these variables.

Then

$$Du = du + \delta u,$$

and

$$du = \frac{dV}{dx_1}dx_1 + \frac{dV}{dx_2}dx_2 + \text{etc.}, = \left(\frac{\partial V}{\partial x_1} + \frac{\partial V}{\partial u_1}\cdot\frac{\partial u_1}{\partial x_1} + \frac{\partial V}{\partial u_2}\cdot\frac{\partial u_2}{\partial x_1} + \cdots\right)dx_1$$
$$+ \left(\frac{\partial V}{\partial x_2} + \frac{\partial V}{\partial u_1}\cdot\frac{\partial u_1}{\partial x_2} + \frac{\partial V}{\partial u_2}\cdot\frac{\partial u_2}{\partial x_2} + \cdots\right)dx_2 + \text{etc.}$$

Now, if the form of the function u_1 vary, while everything else remains constant, then, since V is a function of u_1, the corresponding change in V will be

$$\frac{\partial V}{\partial u_1}\delta u_1,$$

And therefore

$$\delta V = \frac{\partial V}{\partial u_1}\delta u_1 + \frac{\partial V}{\partial u_2}\delta u_2 + \text{etc.}$$

(4) Let $U = F.V$, where V has the same meaning as in (3), and F obeys the distributive law. Then *du* is obtained in the same manner as in (3), viz.,

$$du = \frac{dU}{dx_1}dx_1 + \frac{dU}{dx_2}dx_2 + \text{etc.}$$

In determining the variation, a reader to whom the subject is new might easily fall into the error of assuming the variation arising from a change in the form of u_1 to be

$$\frac{\partial U}{\partial u_1}\delta u_1.$$

It must be borne in mind, however, that the validity of the theorem of the Differential Calculus, from which this conclusion might be erroneously drawn, depends on the assumption that U is a function of u_1. This is, however, not the case, the relation between them being one of form, not one of magnitude. But, since $U = F.V$, we have $\delta U = F.\delta V$, and therefore, from (3),

$$\delta U = F.\left(\frac{\partial V}{\partial u_1}\delta u_1 + \frac{\partial V}{\partial u_2}\delta u_2 + \text{etc.}\right).$$

(5) Let $y=\phi(x)$ be an undetermined function of the single variable x, then, to determine the complete variation of $\frac{d^n y}{dx^n}$, we shall have from (2)

$$d\frac{d^n y}{dx^n}=\frac{d^{n+1}y}{dx^{n+1}}$$

$$\delta\frac{d^n y}{dx^n}=\frac{d^n\delta y}{dx^n},$$

so that

$$D\frac{d^n y}{dx^n}=\frac{d^{n+1}y}{dx^{n+1}}dx+\frac{d^n\delta y}{dx^n},$$

and the function ψ, such that $\delta y=i\psi x$, must be such as to make $\frac{d^n\,.\,\psi x}{dx^n}$ finite for all admissible values of x.

(6) Let $y=\phi(x)$ be an undetermined function of x, and

$$V=f\left(x,\,y,\,\frac{dy}{dx},\,\frac{d^2y}{dx^2}\right)$$

a determined function of the quantities within the brackets. This is a case of (3), putting

$$u_1=y\,,\ \ u_2=\frac{dy}{dx}\,,\ \ u_3=\frac{d^2y}{dx^2}\,.$$

Making these substitutions, writing for brevity

$$M=\frac{\partial V}{\partial x}\,,\ \ N=\frac{\partial V}{\partial y}\,,\ \ P_1=\frac{\partial V}{\partial\frac{dy}{dx}}\,,\ \ P_2=\frac{\partial V}{\partial\frac{d^2y}{dx^2}}\,,$$

and substituting for $\delta\frac{dy}{dx}$ and $\delta\frac{d^2y}{dx^2}$, their values obtained from (5),

$$DV=\left(M+N\frac{dy}{dx}+P_1\frac{d^2y}{dx^2}\right)dx+N\delta y+P_1\frac{d\delta y}{dx}+P_2\frac{d^2\delta y}{dx^2}\,.$$

(7) Let y and V have the same meaning as in (6), and

$$U=\int_{x_0}^{x_1}V\,dx\,.$$

The value of U may then be varied, (1) by a change in the upper limit x_1; (2) by a change in the lower limit x_0; (3) by a change in the form of the function ϕ.

It is unnecessary, as Jellett points out, in investigating the variation of a definite integral, to assign an increment to the independent variable, for the value of U depends only on the values of x_1 and x_0, and on the form of ϕ, and therefore the complete variation cannot contain a term depending on the increment of the independent variable, as distinguished from the increments of its limiting values. We shall find, accordingly, that if an increment be assigned to the independent variable, the coefficient of that increment, under the sign of integration, will become identically equal to zero in the final result.

Let V_1 and V_0 be the values of V at the upper and lower limit, respectively, and let the limits be increased by dx_1 and dx_0, respectively, the form of ϕ remaining unchanged. The respective increments of U will then be V_1dx_1 and $-V_0dx_0$, and therefore

$$DU = V_1dx_1 - V_0dx_0 + \delta\int_{x_0}^{x_1} Vdx.$$

Now, the operation represented by the sign of integration obeys the distributive law, and hence

$$\delta\int_{x_0}^{x_1} Vdx = \int_{x_0}^{x_1} \delta Vdx.$$

Therefore, by (6),

$$\delta V = N\delta y + P_1\frac{d\delta y}{dx} + P_2\frac{d^2\delta y}{dx^2},$$

and therefore

$$\delta\int_{x_0}^{x_1} Vdx = \int_{x_0}^{x_1}\left(N\delta y + P_1\frac{d\delta y}{dx} + P_2\frac{d^2\delta y}{dx^2}\right).$$

The second and third terms within the brackets may be reduced by the method of integration by parts, viz.,

$$\int P_1\frac{d\delta y}{dx}dx = P_1\delta y - \int\frac{dP_1}{dx}\delta y dx,$$

$$\int P_2\frac{d^2\delta y}{dx^2}dx = P_2\frac{d\delta y}{dx} - \int\frac{dP_2}{dx}\cdot\frac{d\delta y}{dx}dx$$

$$= P_2\frac{d\delta y}{dx} - \frac{dP_2}{dx}\delta y + \int\frac{d^2P_2}{dx^2}\delta y dx.$$

Performing these integrations between the limits, and collecting the coefficients of the several quantities, we obtain finally

$$DU = V_1 dx_1 - V_0 dx_0 ,$$

$$+\left(P_1 - \frac{dP_2}{dx}\right)_{x_1}\delta y_1 - \left(P_1 - \frac{dP_2}{dx}\right)_{x_0}\delta y_0 + \left(P_2 \frac{d\delta y}{dx}\right)_{x_1} - \left(P_2 \frac{d\delta y}{dx}\right)_{x_0}$$

$$+\int_{x_0}^{x_1}\left(N - \frac{dP_1}{dx} + \frac{d^2P_2}{dx^2}\right)\delta y dx .$$

The terms in the first line of this expression are independent of any change in the form of ϕ, and depend only on the variation of the limits. The terms in the second line depend on the change in the form of ϕ, but only for the limiting values of x, as shown by the subscripts, which indicate the values to be substituted for x in the general expressions. The term under the sign of integration depends on the general change in the form of the function.

If U is to be made a maximum or a minimum, we must have $DU = 0$, which may be written in the form

$$a_1 - a_0 + \int_{x_0}^{x_1} \beta \delta y dx = 0 ,$$

where β is the expression under the sign of integration, and a_1 and a_0 are the sums of the terms depending on x_1 and x_0 respectively.

Now, the above equation cannot be satisfied, without restricting the generality of δy, unless $a_1 - a_0$ and β vanish separately, for if $a_1 - a_0$ does not vanish, we must have

$$\int_{x_0}^{x_1} \beta \delta y dx = a_0 - a_1 ,$$

an equation which would imply the manifest impossibility that an arbitrary function is expressible in terms of the limiting values of itself and a certain number of its differential coefficients. It follows that $a_1 - a_0$ and $\int_{x_0}^{x_1} \beta \delta y dx$ must vanish separately. Now, the latter expression can only be made to vanish by making $\beta = 0$, since it is inadmissible to restrict the generality of the function δy. That is to say,

$$N - \frac{dP_1}{dx} + \frac{d^2P_2}{dx^2} = 0 ,$$

which therefore determines the form of the function ϕ which will make U a maximum or minimum.

(8) Let

$$U=\int_{x_0}^{x_1} V dx,$$

where

$$V=f\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}, z, \frac{dz}{dx}, \frac{d^2z}{dx^2}\right),$$

y and z being undetermined functions of x.

The complete variation may be determined, as in (7), by adding expressions containing terms N', P'_1, P'_2, etc., depending on the variation of z, exactly as N, P_1, P_2, etc., depend on the variation of y, to the value there obtained for DU.

The extension to functions of any number of undetermined functions of the independent variable, and of its differential coefficients up to any order, is obvious.

(9) Let the functions y and z be connected by a differential or other equation $L=0$. This will not affect the method of obtaining the complete variation, but the variations δy and δz will not then be really independent, and therefore, in order to apply them to the solution of problems, one of them must be eliminated. If the equation $L=0$ can be solved so as to express one of the two quantities y and z as a function of the other, *e.g.* if it can be transformed into such a form as

$$z=\chi\left(x, y, \frac{dy}{dx}, \text{etc.}\right),$$

then the values of $\frac{dz}{dx}$, $\frac{d^2z}{dx^2}$, etc., can be obtained by simple differentiation, and substituted in V, which will then become simply a function of x, y, and its differential coefficients. This is, however, not often possible, and Lagrange has given a general method by which $\delta\int V dx$ may be reduced to a form depending on only one of the variations δy or δz.

Let

$$\alpha=\frac{\partial L}{\partial y}, \quad \beta=\frac{\partial L}{\partial \frac{dy}{dx}}, \quad \gamma=\frac{\partial L}{\partial \frac{d^2y}{dx^2}}, \text{ etc.,}$$

$$\alpha'=\frac{\partial L}{\partial z}, \quad \beta'=\frac{\partial L}{\partial \frac{dz}{dx}}, \quad \gamma'=\frac{\partial L}{\partial \frac{d^2z}{dx^2}}, \text{ etc.}$$

Then

$$\delta L = a\delta y + \beta\frac{d\delta y}{dx} + \gamma\frac{d^2\delta y}{dx^2} + \text{etc.}, + a'\delta z + \beta'\frac{d\delta z}{dx} + \gamma'\frac{d^2\delta z}{dx^2} + \text{etc.},$$

and since the equation $L = 0$ must be satisfied by all admissible forms of the functions y and z, we must have $\delta L = 0$. If this equation can be integrated, so as to express either δy or δz in terms of the other, the required reduction can obviously be effected, but this is not usually possible.

Now, since δL is identically zero, we may write, where λ is an undetermined quantity,

$$\begin{aligned}\delta V &= \delta V + \lambda\delta L \\ &= (N + \lambda a)\delta y + (P_1 + \lambda\beta)\frac{d\delta y}{dx} + \text{etc.}, \\ &+ (N' + \lambda a')\delta z + (P_1 + \lambda\beta')\frac{d\delta z}{dx} + \text{etc.}\end{aligned}$$

We shall then have $\delta U =$ terms depending only on the limits

$$+ \int_{x_0}^{x_1}\left(N + \lambda a - \frac{d(P_1 + \lambda\beta)}{dx} + \text{etc.}\right)\delta y$$

$$+ \int_{x_0}^{x_1}\left(N' + \lambda a' = \frac{d(P_1' + \lambda\beta')}{dx} + \text{etc.}\right)\delta z.$$

If λ is determined by equating the coefficient of δy to zero, this expression will depend only on δz, while, if the coefficient of δz is made to vanish, it will depend only on δy.

(10) Let $u = \phi(x, y, z)$ be an undetermined function of the three independent variables x, y, z, then we shall have, just as in (5),

$$D\frac{d^{m+n+p}u}{dx^m dy^n dz^p} = \frac{d^{m+1+n+p}u}{dx^{m+1}dy^n dz^p} + \frac{d^{m+n+1+p}u}{dx^m dy^{n+1}dz^p} + \frac{d^{m+n+p+1}u}{dx^m dy^n dz^{p+1}} + \frac{d^{m+n+p}\delta u}{dx^m dy^n dz^p}.$$

This result may obviously be extended to any number of variables, as may also the following, in which three independent variables are again taken as an example of the general method.

(11) Let

$$V = f\left(x, y, z, u, \frac{du}{dx}, \frac{du}{dy}, \frac{du}{dz}, \frac{d^2u}{dx^2}, \frac{d^2u}{dxdy}, \ldots\right).$$

Following Jellett, we will write, as suggested to him by the Rev. W. Roberts,

$$\frac{dV}{d\dfrac{d^{m+n+p}u}{dx^m dy^n dz^p}} = Vx^m y^n z^p,$$

Then we find, as in (5),

$$\begin{aligned} DV = {} & \frac{dV}{dx}dx + \frac{dV}{dy}dy + \frac{dV}{dz}dz \\ & + \frac{\partial V}{\partial u}\delta u + V_x \frac{d\delta u}{dx} + V_y \frac{d\delta u}{dy} + V_z \frac{d\delta u}{dz} + V_{xx}\frac{d^2\delta u}{dx^2} + V_{xy}\frac{d^2\delta u}{dxdy} + \text{etc.} \end{aligned}$$

It has already been pointed out that in the mechanical applications of the Calculus of Variations, where the symbol δ is employed

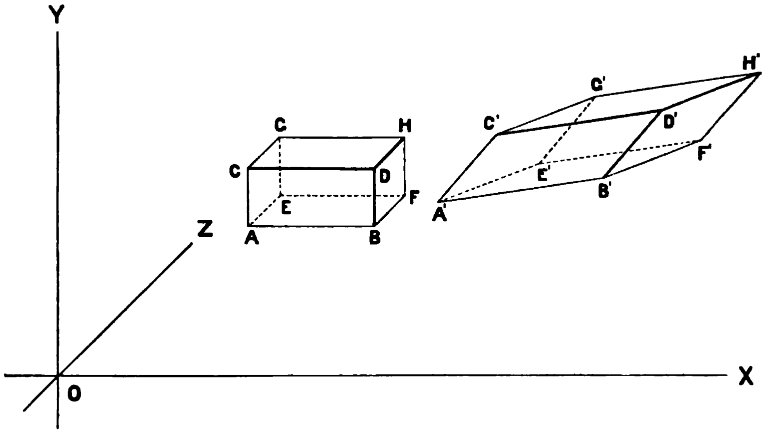


Fig. 32.

to denote a displacement, it ceases to be identical with the symbol d, as applied to an independent variable, the latter having reference to the distance between one particle and another, and the former denoting the displacement of the *same* particle. Let F be a force tending to produce a change $\delta\omega$ in a certain quantity ω, then the moment of the force, as defined by Lagrange, will be $F\delta\omega$. In molecular physics, where we have to deal with forces acting upon every element of a body, and which cannot be adequately defined with respect only to the change produced by their action on a finite portion of the body, we must evidently

take account of the *variation* of an element in the new sense of its change in geometrical magnitude as a result of the displacement of the particles of which the element consists. It therefore becomes necessary to determine the mechanical variation, so arising, of an element of volume, $dxdydz$, of a continuous body. It is of course necessary, in order that mathematical treatment may be possible, that the displacement should follow some regular law determining the relation of the displacement of any particle to its original position. The displacements must therefore be expressible as functions of the co-ordinates. The variation $\delta . dxdydz$ is then obtained by Lagrange as follows[1]:

Let x, y, z be the co-ordinates of a particle A in the body, and let the rectangular parallelepiped AH, having its edges parallel to the axes of co-ordinates, represent the element of volume $dxdydz$. Let A′H′ represent the same element of volume after displacements in its molecules such that the components of the displacement of A are ξ, η, ζ, the displacements of the several molecules being determinable by the expression of each of the displacements ξ, η, ζ as functions of the initial co-ordinates. Let α', β', γ' be the angles between A′B′ and A′E′, A′C′ and A′E′, A′B′ and A′C′ respectively. Then the co-ordinates of the points A′, B′, C′, D′ are:—

Of A′, $x+\xi\,,\ y+\eta\,,\ z+\zeta\,.$

Of B′, $x+dx+\xi+\dfrac{d\xi}{dx}dx\,,\ y+\eta+\dfrac{d\eta}{dx}dx\,,\ z+\zeta+\dfrac{d\zeta}{dx}dx\,.$

Of C′, $x+\xi+\dfrac{d\xi}{dy}dy\,,\ y+dy+\eta+\dfrac{d\eta}{dy}dy\,,\ z+\zeta+\dfrac{d\zeta}{dy}dy\,.$

Of D′, $x+dx+\xi+\dfrac{d\xi}{dx}dx+\dfrac{d\xi}{dy}dy\,,\ y+dy+\eta+\dfrac{d\eta}{dx}dx+\dfrac{d\eta}{dy}dy\,,$

$$z+\zeta+\frac{d\zeta}{dx}dx+\frac{d\zeta}{dy}dy\,.$$

Therefore

$$\mathrm{A'B'}=dx\sqrt{\left\{\left(1+\frac{d\xi}{dx}\right)^2+\left(\frac{d\eta}{dx}\right)^2+\left(\frac{d\zeta}{dx}\right)^2\right\}}=\mathrm{C'D'}.$$

$$\mathrm{A'C'}=dy\sqrt{\left\{\left(\frac{d\xi}{dy}\right)^2+\left(1+\frac{d\eta}{dy}\right)^2+\left(\frac{d\zeta}{dy}\right)^2\right\}}=\mathrm{B'D'}.$$

The force A′B′C′D′ is therefore a parallelogram, and proceeding in the same way with the remaining forces, the element A′H′ is

[1] *Mécanique Analytique*, pp. 191–194.

seen to remain a parallelepiped whose three edges, neglecting quantities of the second order, are

$$\left(1+\frac{d\xi}{dx}\right)dx\,,\ \left(1+\frac{d\eta}{dy}\right)dy\,,\ \left(1+\frac{d\zeta}{dz}\right)dz\,,$$

and their product, as far as terms of the first order, will be

$$\left(1+\frac{d\xi}{dx}+\frac{d\eta}{dy}+\frac{d\zeta}{dz}\right)dxdydz\,.$$

Now,

$$2\mathrm{A'B'}\,.\,\mathrm{A'C'}\cos\gamma' = \mathrm{A'B'}^2+\mathrm{A'C'}^2-\mathrm{B'C'}^2\,,$$

giving by substitution and reduction

$$\cos\gamma' = \frac{d\xi}{dy}+\frac{d\eta}{dx}\,;$$

and similarly

$$\cos\beta' = \frac{d\zeta}{dx}+\frac{d\xi}{dz}\,,$$

$$\cos\gamma' = \frac{d\eta}{dz}+\frac{d\zeta}{dy}\,.$$

These angles therefore differ infinitely little from right angles, and therefore, to the first order, the volume of the new parallelepiped will be represented by the product of its edges, so that the variation of $dxdydz$ will be

$$\left(\frac{d\xi}{dx}+\frac{d\eta}{dy}+\frac{d\zeta}{dz}\right)dxdydz\,.$$

When ξ, η, ζ, instead of being finite, are indefinitely small, or what were formerly called *virtual* displacements, they are usually denoted by the symbol δ, so that

$$\delta\,.\,dxdydz = \left(\frac{d\delta x}{dx}+\frac{d\delta y}{dy}+\frac{d\delta z}{dz}\right)dxdydz\,.$$

(12) To determine the mechanical variation of

$$\mathrm{V} = f\left(x\,,\,y\,,\,z\,,\,\xi\,,\,\eta\,,\,\zeta\,,\,\frac{d\xi}{dx}\,,\,\frac{d\xi}{dy}\,,\,\frac{d\xi}{dz}\,,\,\frac{d\eta}{dx}\,,\ \text{etc.}\right),$$

containing differential coefficients of ξ, η, ζ of any order.

Assuming, as before, that ξ, η, ζ are functions of x, y, z, it is clear that V can only be varied either by a change in some one of the quantities x, y, z, or by a change in the form of some one of the functions ξ, η, ζ. Since the mechanical variation

denotes a change referring throughout to the same particle, increments relating to changes in the distance between two particles in the body are excluded, so that

$$
\begin{aligned}
\delta V = {} & \frac{dV}{dx}\delta x + \frac{dV}{dy}\delta y + \frac{dV}{dz}\delta z \\
& + \frac{\partial V}{\partial \xi}\delta\xi + \frac{\partial V}{\partial \frac{d\xi}{dx}}\,\frac{d\delta\xi}{dx} + \text{etc.} \\
& + \frac{\partial V}{\partial \eta}\delta\eta + \frac{\partial V}{\partial \frac{d\eta}{dx}}\,\frac{d\delta\eta}{dx} + \text{etc.} \\
& + \frac{\partial V}{\partial \zeta}\delta\zeta + \frac{dV}{\partial \frac{d\zeta}{dx}}\,\frac{d\delta\zeta}{dx} + \text{etc.},
\end{aligned}
$$

which is simply the mathematical variation, with the addition of the terms in the first line, involving the quantities δx, δy, δz, which cannot occur in the latter.

APPENDIX B.

HAMILTON'S PRINCIPLE.

THE most formidable of the objections which have been urged against Hamilton's principle as a fundamental principle of mechanics is that taken by Hertz—viz. that it does not apply to all natural motions, for it requires that, if fixed connections exist between the chosen co-ordinates of a system, these connections must be expressible mathematically by finite equations between the co-ordinates; in other words, the configuration of the system at any instant must be expressible in terms of the co-ordinates and their first differential coefficients with regard to the time, all these quantities being capable of independent variation. Hertz calls such systems *holonomic.* This is not always the case in natural motions: for example, when bodies of three dimensions roll upon each other without slipping, there must necessarily be at least one non-integrable differential equation connecting the co-ordinates.

Sir Joseph Larmor attempts to meet this objection by asserting that "it may be urged that the motion of rolling is foreign to molecular dynamics, on which the laws of mechanical dynamics must be ultimately based."[1] It might, on the other hand, be urged that, considering how very little we really know about the nature of molecular mechanism, we have no right to categorically exclude even this notion of rolling; and if Larmor's contention be admitted, it does not answer Hertz's objection. To do this we should have to maintain that no motions can occur in molecular dynamics of a kind which would lead to non-integrable differential equations between the co-ordinates, and I think this position can be maintained when we are dealing with the ultimate structure of the substance of the universe. Jeans observes[2] that in such investigations it is almost unthinkable that we should be concerned with a non-holonomic system, since, except in text-

[1] *Æther and Matter*, p. 277.

[2] "General Dynamics of Matter and Ether," *Proc. Roy. Soc.*, vol. lxxvi., A, 1905, p. 276.

books, such systems can occur only through the massing together of a large number of independent co-ordinates into a smaller number, which nearly, but not quite, represent the larger true number. Even in dealing with problems of a less fundamental character, Hamilton's principle has proved to be a trustworthy guide over so wide a range of natural phenomena, that if motions of the kind to which it does not apply do occur in nature, they can only do so very rarely, and the balance of probability is, therefore, in any special case, largely in favour of its employment being legitimate. The limitations set forth on page 51 must, of course, always be borne in mind, and in cases where the relative motions of the parts of a system introduce a passive reaction of frictional type, the nature and amount of these reactions will have to be elucidated as a preliminary to the application of the principle of Least Action. In such cases it will generally be found that a mechanical theory is possible only as a first approximation in problems where, the viscous forces being small, they may be considered as being proportional to their originating circumstances. Sir Joseph Larmor[1] cites the ordinary statical theory of elasticity as an example of this, observing that "a theory of transmission of stress in elastic matter exists in the ordinary sense, which considers the action of the surrounding medium on any given element of it as constituted solely of tractions over its surface, only because the range of the molecular forces of cohesion is very small compared with the dimensions of an element of volume which can be effectively treated as infinitesimal."

I was inclined for a time to seek in Hertz's suggestion of possible exceptions to the validity of Hamilton's principle a solution of the discrepancy between the observed ratio of the two specific heats of a gas and the ratio calculated from the Boltzmann-Maxwell law of the equipartition of energy between the different degrees of freedom of the gaseous molecules, and suggested its possibility in a review of Sir Joseph Larmor's *Æther and Matter* in *Feilden's Magazine* for September 1902, from which part of this appendix is reproduced. In a letter which I received subsequently from Lord Kelvin he expressed his absolute conviction of the universal validity of Hamilton's principle, subject, of course, to the necessary limitations which have been indicated.

The discrepancy was soon afterwards completely accounted for by Professor J. H. Jeans in his *Dynamical Theory of Gases*, which was published in 1904. In this work he gave a new proof of the equipartition theory which appears to me to be absolutely

[1] *Æther and Matter*, p. 270.

conclusive, under the limitations which he showed to be necessary, viz. that the conditions were such that the interaction between ether and matter could be treated as negligible. When this condition is not fulfilled the theorem does not hold, and the discrepancy previously referred to was a consequence of the illegitimate application of the theorem.

Another objection which has been raised against the principle of least action is that it cannot be fundamental, inasmuch as it determines the present course of a system by a reference to its future as well as to its past.

Sir Joseph Larmor's reply to this is that

"the objection will be removed if we bear in mind that the complete system is of very complex molecular constitution, and that the principle of action is really only an algorithm, constructed so as to enable us to abstract the molecular details, while retaining all that relates to the matter in bulk."

An objection of a somewhat similar character might be raised against other modern analytical methods, such as that of the Vector Potential, inasmuch as by the procedure of integration through all space they make the state of any system at any given time depend on the simultaneously existing conditions throughout the universe.

The reply which I should make to both objections would be that, in using such general methods of dynamical reasoning, we contemplate whatever system may be under consideration as part of the sum-total of things, or the universe, which we can only think of as being extended both in space and time. Of this sum-total of things we endeavour to abstract so much as will determine the system to the extent required, if possible, and if not, as far as we are able, and ignore the remainder. From the philosophical point of view, the space and time relations to our system of the utilised portion of the sum-total are a matter of indifference. An argument which further illustrates and enforces this point of view will be found in the preface. It occurred to me too late for insertion here.

It may also be pointed out that the determination of the present state of a system by its future as well as its past is paralleled in everyday life whenever our actions are determined by future events.

APPENDIX C.

MAXWELL'S ANALYTICAL METHODS.

THE study of Maxwell's treatise will be greatly facilitated by always bearing in mind that he makes extensive use of two mathematical methods, very distinct in themselves, and that he not infrequently passes from one to the other without clearly marking the lines of demarcation. When this is kept in mind, such passages from one method to the other will usually be sufficiently obvious to prevent any confusion, but if the reader is not on his guard in this respect, they considerably enhance the inherent difficulties of the subject. These two methods are clearly distinguished in the following paragraph, which forms number 529 of Maxwell's treatise:—

"We are accustomed to consider the universe as made up of parts, and mathematicians usually begin by considering a single particle, and then conceiving its relation to another particle, and so on. This has generally been supposed the most natural method. To conceive of a particle, however, requires a process of abstraction, since all our perceptions are related to extended bodies, so that the idea of the *all* that is in our consciousness at a given instant is perhaps as primitive an idea as that of any individual thing. Hence there may be a mathematical method in which we proceed from the whole to the parts instead of from the parts to the whole. For example, Euclid, in his first book, conceives a line as traced out by a point, a surface as swept out by a line, and a solid as generated by a surface. But he also defines a surface as the boundary of a solid, a line as the edge of a surface, and a point as the extremity of a line.

"In like manner we may conceive the potential of a material system as a function found by a certain process of integration with respect to the masses of the bodies of the field, or we may suppose these masses themselves to have no other mathematical meaning than the volume-integrals of $\frac{1}{4\pi}\nabla^2 V$, where V is the potential.

"In electrical investigations we may use formulæ where the quantities involved are the distances of certain bodies, and the electrifications or currents in these bodies; or we may use formulæ which involve other quantities, each of which is continuous through all space.

"The mathematical process employed in the first method is integration along lines, over surfaces, and throughout finite spaces; those employed in the second method are partial differential equations and integrations throughout all space.

"The method of Faraday seems to be intimately related to the second of these modes of treatment. He never considers bodies as existing with nothing between them but their distance, and acting on one another according to some function of that distance. He conceives all space as a field of force, the lines of force being in general curved, and those due to any body extending from it on all sides, their directions being modified by the presence of other bodies. He even speaks of the lines of force belonging to a body as in some sense part of itself, so that in its action on distant bodies it cannot be said to act where it is not. This, however, is not a dominant idea with Faraday.

"I think he would rather have said that the field of space is full of lines of force, whose arrangement depends on that of bodies in the field, and that the mechanical and electrical action on each body is determined by the lines which abut on it."

There is a similar choice of methods to be employed in the development of the electron theory, and as Sir Joseph Larmor's development of this theory is dealt with in Chapter IX., it will be an assistance to the mathematical reader to have before him Larmor's remarks on these two methods, as given in his papers on "A Dynamical Theory of the Electric and Luminiferous Medium" (*Philosophical Transactions of the Royal Society*, 1895, A, p. 705).

"We may start from the dynamical equations of the free ether, which are known, and which apply exactly to a sub-element of volume which does not contain any electrons; we can make use of the various forces and fluxes which have proved useful in forming a representation of electrical phenomena, and by their help we can replace these dynamical equations, which are of the second order in differentiations, by the two circuital relations which are each of the first order. The advantages, both for analysis and for explanation, of a transformation of this kind, have long been recognised in pure dynamics. Now we can pass, by integration of these relations, from the sub-element of volume which contains no electrons, to the effective element of volume which contains a number of electrons large enough to enable us to smooth out their individual peculiarities and so retain a continuous differential analysis. The transformation is most conveniently effected so that the circuital relations shall remain unaltered in form, while the meanings of the various quantities that enter into them are modified by taking account of the presence of the distribution of electrons in the enlarged element of volume. This method of smoothing out the molecular discreteness of the ether is the one that has been most in favour in recent years; but when we come to the consideration of moving media and convection currents, and particularly of ponderomotive forces, it will be necessary to supplement it by the more direct procedure now to be described.

"The other method of analysis involves the utilisation of the ideas of

the electrodynamic potential in an amended form. It is concerned primarily with the dynamics of a single electron in an electric field given as regards motion and strain, or kinetic and potential energy; it forms the kinetic energy T, and the potential energy W, of the electron, and thence its fundamental Lagrangian function or electrodynamic potential $T - W$. The relations between the motions and the forcives of the electron are now to be deduced by dynamical methods; and there only remains a process of summation in order to determine the equations that are appropriate to the various modes of co-ordinated groupings and movements of electrons that have been enumerated above. This process has a fundamental theoretical superiority over the previous one, in that it includes in its scope the ponderomotive forces which act on the bodies in the field. The various discussions in Maxwell's 'treatise,' which proceed by use of the vector potential of electrodynamic induction, are to be classed as essentially related to this point of view."

APPENDIX D.

MAXWELL'S ETHER STRESS SYSTEMS.

THE electric stress considered by Maxwell in §§ 105 to 110 of his treatise, and which is the one referred to on pages 117 and 122, may be regarded as a special case, applicable only to free ether, of the more general stress-system dealt with on page 155. It leads, in the case of homogeneous media, to interfacial tractions only, without bodily force; and therefore, as Sir Joseph Larmor points out, it can only be validly applied to unpolarised media, and he observes that:—

"The proposition really established is that the mechanical forcive due to attraction at a distance, obeying the law of inverse squares, between material bodies, may be represented by a connection in the form of an imposed extraneous stress symmetrical with respect to the lines of force, acting across the intervening medium, *provided* that medium is not in any way polarised by the force. A stress restricted by this relation of symmetry involves only two variables, the principal tractions along and at right angles to the line of force; and the essence of Maxwell's theorem is that it is possible always to determine these two variables so as to satisfy the three equations of equilibrium of the element of volume of the medium." . . . . "The proposition is in itself so remarkable that it deserves to be formulated abstractedly without reference to hypothetical applications."[1]

Both this simple stress-system applicable only to ether free from electrons, and the more general system considered on page 155, which is applicable to polarisable media, are often extremely convenient for purposes of calculation. It must, however, always be borne in mind that they must be regarded merely as crude models, and not as representing actual distributions of stress in the ether. They are models of great value, but, like all such models, they may lead us into error if their limitations are overlooked. When they are regarded as representing actual distributions of stress in the ether they are seen immediately to involve a fundamental difficulty. We know that the electromagnetic properties of the

[1] *Phil. Trans.*, vol. cxc., A, 1897, p. 254.

ether are expressible by a system of linear equations as a consequence of the observed fact that the velocity of propagation of electromagnetic disturbances is independent both of the wavelength and the amplitude. We should therefore expect any actual distribution of stress to follow a linear law, whereas the Maxwell stresses, instead of being proportional to the electric force, are proportional to its square.

It is pointed out by Dr C. V. Burton[1] that it follows as a consequence of the principle of conservation of energy that if the ordinary theory of stresses is to be applied, as is done by Maxwell, to the case of a conductor surrounded by a dielectric, the surface forming the boundary of the conductor must be regarded as an impermeable barrier between it and the dielectric medium, so that if the surface of the conductor be supposed to advance in the direction of the outwardly drawn normal, the dielectric medium must recede to the same extent, and *vice versa*. There must not therefore be any interpenetration, even in the case of a conductor surrounded by free ether. Under these circumstances, a given distribution of stress in the dielectric medium will give rise to determinate forces acting upon the conductor, but this would no longer be the case if the conductor were at all permeable by the dielectric, as we know to be the case when this consists of ether. The conducting body would have to be treated as foreign to, and wholly distinct from, the ether.

Dr Burton shows, moreover, that this assumption necessarily leads to the conclusion that the ether must be indefinitely compressible. For consider a condenser consisting of a conducting sphere of radius r_1, contained in a concentric spherical cavity of radius r_2 in an outer conducting body, the interspace containing only ether. Let e be the charge on the inner sphere, then $-e$ will be the charge on the hollow spherical surface. Passing through the interspace in the direction of the outwardly drawn radius, the electric force will have values ranging from e/r_1^2 to e/r_2^2. Now, the inner sphere being insulated, so that its charge remains constant, let its radius be increased by dr_1, then the increase in the electrostatic energy of the system will be

$$\frac{d}{dr_1}\left\{\frac{1}{2}e^2\left(\frac{1}{r_1}-\frac{1}{r_2}\right)\right\}dr_1 = -\frac{1}{2}\,\frac{e^2}{r_1^2}dr_1, \qquad . \qquad . \quad (1)$$

while the Maxwell tension per unit area of the inner sphere is

$$\frac{1}{8\pi}\left(\frac{e}{r_1^2}\right)^2,$$

[1] *Phil. Mag.*, vol. xvii., 1909, p. 641.

and the work per unit area due to it is

$$\frac{1}{8\pi}\frac{e^2}{r_1^4}dr_1.$$

The total work done by the ethereal mechanical stress during the expansion of the sphere is therefore

$$\frac{1}{2}\frac{e^2}{r_1^2}dr_1,$$

which is simply the expression on the right-hand side of (1) with its sign changed, that is to say, it is the *loss* of electrostatic energy of the system.

Proceeding in this way, the thickness $r_2 - r_1$ of the interspace may be reduced from an original value as large as we please to an insignificant amount, the contained ether always retaining a radial tension which is nowhere less than e/r_2^2, and being made to occupy ultimately only a minute fraction of its original volume.

Dr Burton then compares the interpretations of the small energy variation corresponding to a small change in the configuration of an electrified system; (1) in terms of Maxwell's mechanical stress system; and (2) in terms of the electronic theory.

(1) "Let the system consist of any number of bodies carrying surface charges, the intervening spaces, for simplicity of illustration, being supposed vacuous; and consider the virtual work of the electrostatic forces when the system suffers a determinate change of configuration $\delta\theta$, specifiable in terms of the geometrical co-ordinates. If, without making any other change, we increase the charge of every element until it is n times as great as before, the virtual work corresponding to $\delta\theta$ will be n^2 times as great as before. Now, on the Faraday-Maxwell view, since we regard the ether as simply partaking of the normal displacement of each electrified surface, the displacement of the ether involved in the change of configuration $\delta\theta$ will depend on $\delta\theta$ only, and not on the electrical charges. Hence, when all the charges are raised to n-fold, involving an n^2-fold value for the virtual work in $\delta\theta$, we must have n^2-fold values for the ethereal stresses."

(2) "On the electronic theory, any displacement $\delta\theta$ of the bodies in question is a displacement of the electrons which make up those bodies and their surface charges, and the corresponding displacement of the ether at any point is the vector increment of electric polarisation at that point. On this view, then, the n^2-fold virtual work corresponding to $\delta\theta$, due to n-fold charges throughout the system, is accounted for by n-fold virtual displacement everywhere in the ether surrounding the bodies, together with n-fold ethereal stress."

The ether stress is therefore, according to the electronic theory, simply proportional to the electric force, which is in conformity with the linearity of the electromagnetic equations.

Exactly similar considerations apply to the generalised stress-system considered on page 155.

APPENDIX E.

APPLICATION OF GREEN'S PRINCIPLE OF THE EXISTENCE OF A "POTENTIAL" TO THE DERIVATION OF ELECTRIC AND MAGNETIC FORCES.

The principle was enunciated by Green in his paper, "On the Reflection and Refraction of Light at the Common Surface of two Non-crystallised Media," in 1837, in the following terms:—

"In whatever manner the elements of any material system may act upon each other, if all the internal forces be multiplied by the elements of their respective directions, the total sum for any assigned portion of the mass will always be the exact differential of some function."

We may illustrate the value of this principle by applying it to derive the general expression of the resultant electric force and its components in terms of the potential. If s be the length of an arc AB of a curve measured from A, and if the resultant electric force E at each point make an angle ϵ with the tangent drawn in the positive direction, then the work done on a unit quantity of electricity in moving it along an element of the curve will be $\mathsf{E}\cos\epsilon ds$, and the total E.M.F. will be V, where $V = \int_A^B \mathsf{E}\cos\epsilon ds$.

If X, Y, Z are the components of E, this becomes

$$V = \int_A^B \left(X\frac{dx}{ds} + Y\frac{dy}{ds} + Z\frac{dz}{ds}\right)ds.$$

Now $Xdx + Ydy + Zdz$ must be a perfect differential of a function of x, y, z, so that

$$V = \int_A^B (Xdx + Ydy + Zdz) = V_A - V_B,$$

where the integration is performed in any way from the point A to the point B, whether along the given curve or along any other

line between A and B. V is then a scalar function of the position of a point in space, so that when its co-ordinates are known, its position is definitely determined. V is then what was defined in Chapter I. as the potential of the point.

Since the total E.M.F. along the arc AB is $V_A - V_B$, if we denote this arc by ds we shall have $\mathsf{E}\cos\epsilon = -\dfrac{dV}{ds}$, and hence, by assuming ds parallel to each of the axes in succession,

$$X = -\frac{dV}{dx}, \quad Y = -\frac{dV}{dy}, \quad Z = -\frac{dV}{dz},$$

and

$$\mathsf{E} = \left\{ \left(\frac{dV}{dx}\right)^2 + \left(\frac{dV}{dy}\right)^2 + \left(\frac{dV}{dz}\right)^2 \right\}^{\frac{1}{2}}.$$

Similarly, the line integral of the magnetic force H, having the components α, β, γ, along an arc AB, or the magnetomotive force, as it is sometimes called, will be given by the equation

$$\int_A^B (\alpha dx + \beta dy + \gamma dz) = \Omega_A - \Omega_B.$$

Also,

$$\mathsf{H} = \left\{ \left(\frac{d\Omega}{dx}\right)^2 + \left(\frac{d\Omega}{dy}\right)^2 + \left(\frac{d\Omega}{dz}\right)^2 \right\}^{\frac{1}{2}}.$$

APPENDIX F.

DETERMINATION OF POTENTIAL GRADIENTS IN THE INTERIOR OF MAGNETS AND OTHER BODIES.

THE ordinary argument for the necessity of imagining a cavity to be formed at a point in the interior of a magnetised body at which it is desired to determine the gradient of magnetic potential may be stated as follows:—

The aggregate potential V, due to the body, at an external point ξ, η, ζ, will be determined by the equation

$$V = \int (Al + Bm + Cn)\frac{dv}{r^2},$$

where l, m, n are the direction cosines of the vector r drawn to the point ξ, η, ζ, from the element of volume dv, situated at the point x, y, z, at which the components of I, the intensity of magnetisation, are A, B, C. Hence

$$V = \int \left(A\frac{d}{dx} + B\frac{d}{dy} + C\frac{d}{dz}\right)\frac{1}{r}dv.$$

Since ξ, η, ζ represents an external point, no element of this integral can become infinite, and therefore it can be legitimately transformed by integration by parts, so that

$$V = \int (\lambda A + \mu B + \nu C)\frac{1}{r}dS - \int \left(\frac{dA}{dx} + \frac{dB}{dy} + \frac{dC}{dz}\right)\frac{1}{r}dv,$$

where λ, μ, ν are the direction cosines of the normal component of the polarity measured outwards, at the surface element dS of the boundary; that is to say, the potential at an external point, due to the polarised mass, may be represented by that of a volume density

$$\rho = -\left(\frac{dA}{dx} + \frac{dB}{dy} + \frac{dC}{dz}\right),$$

and a surface density σ, equal to the normal component of the polarity at the boundary, measured outwards. The statical

theory of the polarity is in this manner reduced to the simpler theory of continuous distributions of attracting matter.

At an internal point $1/r$ will be infinite for an element of the integral, so that the process of integration by parts would appear in this case to be illegitimate. If this were really the case it would involve an essential difference in the calculation of the potential and its gradient at points within magnetic matter; and at points within a polarised dielectric, or within gravitating matter, and in the two latter cases it has not usually been considered necessary to imagine cavities to be formed. The existence of such a distinction is, however, excluded by the consideration that a magnetic distribution can always be represented alternatively, either as an assemblage of infinitely small doublets, or as a superposition of positive and negative magnetic matter throughout the body, for each of which a separate calculation may be made.

Professor Minchin points out that in calculating the gravitational attraction at a point within a material body, if m and m' are two elementary masses separated by a distance r, the elemental attraction mm'/r^2 does not necessarily become infinite when r vanishes, because the element of mass vanishes at the same time, as is recognised in the mathematical process in which we take

$$m = \rho r^2 \sin\theta dr d\theta d\phi,$$

in ordinary polar co-ordinates. In this case neither m/r nor m/r^2 becomes infinite.

Now the element of magnetic potential at a point P, due to magnetised matter at P, is

$$\text{I}\frac{d\frac{1}{r}}{ds}dv,$$

so that if we take as the element dv, $r^2 \sin\theta dr d\theta d\phi$, instead of $dxdydz$, no infinite element will be introduced into the integral, even when r vanishes.

It follows, therefore, that the magnetic force at a point ξ, η, ζ, whether this point be inside or outside magnetic matter, has the components

$$-\frac{d}{d\xi}\int\left(\text{A}\frac{d}{dx}+\text{B}\frac{d}{dy}+\text{C}\frac{d}{dz}\right)\frac{1}{r}\,dv,$$

$$-\frac{d}{d\eta}\int\left(\text{A}\frac{d}{dx}+\text{B}\frac{d}{dy}+\text{C}\frac{d}{dz}\right)\frac{1}{r}\,dv,$$

$$-\frac{d}{d\zeta}\int\left(\text{A}\frac{d}{dx}+\text{B}\frac{d}{dy}+\text{C}\frac{d}{dz}\right)\frac{1}{r}\,dv.$$

These considerations prove definitely that there is no essential distinction, in the mathematical development of the theory of the potential, between polarised and unpolarised material media. An investigation of the validity of the assumptions on which this development depends will, however, be instructive, and will show the necessity of considering the force within a small cavity in both kinds of media.

We saw in Chapter III. that the legitimacy of applying the processes of continuous integration to the assemblages of discrete molecules which constitute material media depends upon the assumption that the distribution of density and other properties is entirely continuous. This cannot be rigidly true of any discrete medium, and the assumption amounts to replacing the actual discrete distribution by a model in which the distribution is continuous. We have therefore to consider the effects of this substitution and to inquire how far the mathematical results obtained from the model will represent true conclusions with regard to actually existing media. The integrals in the case of the model distributions correspond to actual summations of contributions from very small, but finite, molecules of complex structure. For example, the gravitational potential of the system of molecules constituting a material body will not be an analytical function as regards its second, and higher differential coefficients; for the value of one of these at a point inside the body will depend on the immediately adjacent distribution of the molecules. In the model this potential is replaced by an entirely analytical function, quantitatively equivalent to it as far as the function itself and its first derivatives are concerned. It is this analytical function which is defined by the characteristic equation

$$\nabla^2 V + 4\pi\rho = 0,$$

in which ρ represents the smoothed-out, or average, value of the density of the molecular distribution, and its derivatives of all orders have definite values both inside and outside the body.

As a magnetic illustration Sir Joseph Larmor[1] considers a uniform assemblage of very small magnetised spheres. If the number of spheres per unit volume be increased indefinitely, and their size correspondingly diminished, while the intensity of magnetisation remains constant for each, the magnetic potential of the system at points in the spaces between them may ultimately be represented by the ordinates of a curve which

[1] *Æther and Matter*, p. 260.

may be considered as compounded of a smooth curve on which is superposed an undulation of indefinitely small amplitude and wave-length, each of about the order of the distances between the spheres. The gradient of the amplitude of the fluctuating portion of the function represented by the curve which is superposed on the mean gradient will remain finite, although its wave-length continues indefinitely small. Since molecular magnitudes are finite, the mathematical process cannot be continued to a limit, and therefore a mechanical theory of the properties of the medium will be impossible unless the unknown details of the molecular distribution can be eliminated. The substitution of the model referred to above for the actual distribution corresponds to defining the function represented by the smooth curve as the analytical potential, and its gradient as the magnetic force. Now, proceeding in the actual case as far as possible towards the limit, the minute and practically indeterminable local irregularities disappear from the potential, but only become more marked in its gradient. The substitution will, therefore, be legitimate provided only that these indeterminable minute irregularities in the magnetic force mutually compensate one another's effects throughout the mass when considered in bulk. Its legitimacy therefore follows from the principle of the mutual compensation of molecular forcives considered in Chapter X., where it is shown to be a necessary consequence of the observed permanence in the constitution of material bodies. If it were not so, the molecular structure would be capable of alteration, and might even have its stability destroyed by mechanical causes, so that it would no longer be possible to assume in ordinary mechanical theory that the physical properties of a medium are unaffected by small forces.

Considering the question mathematically, we see that if the actual summations in the interior of bodies are to be replaced by the completely defined integrals of purely mathematical rational dynamics, we must take the principal values, in Cauchy's sense, of these integrals, this being the only method of obtaining determinate values. Such a principal value, however, is simply the value at the centre of a minute spherical cavity.

Similar considerations apply to the determination of the vector potential at points inside magnetic matter, and are discussed by Sir Joseph Larmor in *Æther and Matter*, Appendix A, to which the reader may refer with advantage.

We now see that the integration by parts of the expression is perfectly legitimate at internal as well as external points, so that

the magnetic force **H** is the gradient of a potential due to a volume density ρ and a surface density σ, as in the ordinary theory of continuous distributions of gravitating matter.

Let a, b, c be the components of **B**; α, β, γ those of **H**; and A, B, C those of **I**; then it follows, by Poisson's extension of Laplace's characteristic equation, that everywhere except at a surface of sudden transition,

$$\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}=4\pi\rho,$$

where

$$\rho=-\left(\frac{dA}{dx}+\frac{dB}{dy}+\frac{dC}{dz}\right).$$

Therefore, except at such a surface, the components of **B** satisfy the solenoidal condition

$$\frac{da}{dx}+\frac{db}{dy}+\frac{dc}{dz}=0.$$

Professor Minchin has kindly supplied me with the following proof that the normal flux of **B** through the bounding surface

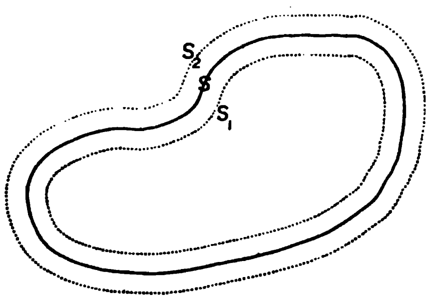


Fig. 33.

of the body is also zero, to replace the one given by Maxwell, which is defective for the reason stated on page 136.

Let S, fig. 33, be the boundary surface, and let the surfaces S_1 and S_2 be described just within, and just without, S, respectively. Let m_1 be the amount of free magnetism within S_1, and let m_0 be the amount on S_1. Then the normal flux of **H** through S_1 is $4\pi m_1$. Also, if ν is the outwardly drawn normal component of **I**, the normal flux of **I** through the surface S_1 is

$$\int \nu dS_1 = m_0.$$

But

$$m_0 = -m_1,$$

so that

$$\text{normal flux of } \mathbf{H} + 4\pi\mathbf{I} = 4\pi m_1 + 4\pi m_0 = 0.$$

Now the normal flux of $\mathbf{H}$ across S_2 is null, and $\mathbf{I}$ is null at every point on S_2, so that the normal flux of $\mathbf{H}+4\pi\mathbf{I}$, that is to say, the normal flux of $\mathbf{B}$, is therefore continuously null in passing through the surface of the body, although it would appear to be indeterminate on separate investigation of the surface itself. $\mathbf{B}$ therefore satisfies the solenoidal condition everywhere, or is everywhere a stream vector.

APPENDIX G.

RELATION BETWEEN THE DI-ELECTRIC CONSTANT AND THE DENSITY OF A FLUID MEDIUM.

THE value $\frac{4}{3}\pi$, which was stated on page 132 to be the value of p, is derived by Sir Joseph Larmor as follows[1]:—

The total electric force on a molecule is derived from the potential

$$V = \sum\left(m_x\frac{d}{dx} + m_y\frac{d}{dy} + m_z\frac{d}{dz}\right)\frac{1}{r},$$

where m_x, m_y, m_z are the components of the moment m of a polarised molecule. For a point within the polarised medium, this involves the actual distribution of neighbouring molecules, and therefore, at any instant, the derived electric force changes rapidly in the molecular interstices. For an outside point, or within a cavity which is large as compared with molecular distances, the summation may be replaced by continuous integration, so that, where f', g', h' are the components of I′,

$$V = \int\left(f'\frac{d}{dx} + g'\frac{d}{dy} + h'\frac{d}{dz}\right)\frac{d\tau}{r},$$

and the derived force is perfectly continuous. Integrating by parts (the origin being outside the region of integration, the function does not become infinite therein),

$$V = \int(lf' + mg' + nh')\frac{dS}{r} - \int\left(\frac{df'}{dx} + \frac{dg'}{dy} + \frac{dh'}{dz}\right)\frac{d\tau}{r};$$

that is to say, the potential at points in free ether is that due to Poisson's ideal distribution of volume density ρ and surface density σ where

$$\rho = -\left(\frac{df'}{dx} + \frac{dg'}{dy} + \frac{dh'}{dz}\right)$$

$$\sigma = lf' + mg' + nh',$$

[1] *Phil. Trans.*, vol. cxc., A, 1907, p. 233.

the surface density being extended over the surface of the cavity for a point within one, as well as over the outer boundary.

Now, in fluids at any rate, the polar molecules are in rapid movement such that the force is one involving an average distribution, and therefore properly regarded as due to a Poisson continuous density, even as regards volume elements close to the point considered.

To find the average force causing polarisation of a molecule, that molecule must be regarded as at the centre of a spherical cavity whose radius is of the order of molecular distances, and account has to be taken of a Poisson distribution on its interior surface; or, more precisely, of the result of an averaged continuous local polarisation, surrounding the molecule, whose intensity increases from zero at a certain distance up to the full amount I' at the limit of the molecular range, this intensity being uniform in direction and a function of the distance only. The force due to this is $\frac{4}{3}\pi I'$ along the direction of the polarisation I', which is therefore the local part to be added on to the electric force as ordinarily defined, viz. to that arising from the density ρ throughout the medium and the density σ on its external surface, and so everywhere derivable from a potential. Thus p should have the value $\frac{4}{3}\pi$ for a fluid, though it may differ somewhat in the case of solids, especially if crystalline. Larmor suggests that the fact that the values of the refractive index for liquids are slightly in excess of what Lorentz's formula would give by computation from the values for their vapours, may be an indication that this averaged field of molecular action is slightly elongated instead of spherical.

APPENDIX H.

THE ALTERNATIVE MECHANICAL REPRESENTATIONS OF ELECTRIC DISPLACEMENT AND MAGNETIC INDUCTION, AND THE MOBILITY OF ELECTRONS.

THE fundamental equations of the electromagnetic field, expressed in terms of electromagnetic units, are

$$\text{curl } \mathbf{H} = 4\pi \mathbf{C}\,; \quad \text{curl } \mathbf{E} = -\dot{\mathbf{B}}\,;$$

or

$$\int \mathbf{H} ds = 4\pi \int \mathbf{C} dS\,; \qquad \int \mathbf{E} ds = -\frac{d}{dt}\int \mathbf{B} dS\,.$$

The correspondence between the two alternative mechanical interpretations, or representations, referred to on page 161, is somewhat obscured by the occurrence of the factor 4π. I shall therefore here follow Lorentz[1] in expressing these equations in terms of units due to Heaviside.

We saw on page 30 that the electromagnetic system was obtained by taking $B = 1$ in free ether or air in the equations representing the repulsions of equal electric charges and equal magnetic poles,

$$A\frac{e^2}{r^2} = B\frac{m^2}{r^2} = F\,.$$

It was pointed out by Heaviside that the factor 4π could be eliminated from the fundamental equations by employing a system of units defined by the condition that in free ether or air

$$A = B = \frac{1}{4\pi}\,.$$

These equations then assume the form

$$\text{curl } \mathbf{H} = \frac{1}{c}\mathbf{C}\,; \quad \text{curl } \mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}\,;$$

[1] *Encyclopädie der mathematischen Wissenschaften*, vol. V2, pp. 63–280.

or

$$\int \mathbf{H}ds = \frac{1}{c}\int \mathbf{C}dS\,; \quad \int \mathrm{E}ds = -\frac{1}{c}\frac{d}{dt}\int \mathbf{B}dS$$

where c is the velocity of radiation.

This system must not be confounded with Heaviside's "Rational Units," in which he takes

$$A = \frac{1}{4\pi\epsilon_0}, \quad B = \frac{1}{4\pi\mu_0}, \quad \epsilon_0\mu_0 = \frac{1}{c^2}.$$

In the latter system the fundamental equations would become

$$\text{curl } \mathbf{H} = \mathbf{C}\,; \quad \text{curl } \mathbf{E} = -\dot{\mathbf{B}}\,;$$

or

$$\int \mathbf{H}ds = \int \mathbf{C}dS\,; \quad \int \mathbf{E}ds = -\frac{d}{dt}\int \mathbf{B}dS.$$

In free ether we shall then have

$$\mathbf{E} = \mathbf{D}, \quad \mathbf{B} = \mathbf{H}, \quad \mathbf{C} = \dot{\mathbf{D}}\,;$$

and

$$\text{div } \mathbf{H} = \frac{\partial \mathbf{H}_x}{\partial x} + \frac{\partial \mathbf{H}_y}{\partial y} + \frac{\partial \mathbf{H}_z}{\partial z} = 0,$$

$$\text{div } \mathbf{D} = \frac{\partial \mathbf{D}_x}{\partial x} + \frac{\partial \mathbf{D}_y}{\partial y} + \frac{\partial \mathbf{D}_z}{\partial z} = 0.$$

The fundamental equations therefore take the form

$$\dot{\mathbf{D}} = c \text{ curl } \mathbf{H}, \quad \dot{\mathbf{H}} = -c \text{ curl } \mathbf{D},$$

subject to the conditions

$$\text{div } \mathbf{D} = 0, \quad \text{div } \mathbf{H} = 0.$$

Now, as on page 161, let

$$\mathbf{u} = \mathrm{A}\mathbf{D}$$

where A is a constant.

Then we have

$$\mathrm{W} = \tfrac{1}{2}k\mathbf{u}^2 = \tfrac{1}{2}k\mathrm{A}^2\mathbf{D}^2, \quad \mathrm{T} = \tfrac{1}{2}\rho\dot{\mathbf{q}}^2 = 2\rho\mathrm{A}^2c^2\mathbf{H}^2.$$

For

$$\mathrm{A}\dot{\mathbf{D}} = \dot{\mathbf{u}} = \tfrac{1}{2} \text{ curl } \dot{\mathbf{q}}, \quad \text{and} \quad \mathrm{A}\dot{\mathbf{D}} = \mathrm{A}c \text{ curl } \mathbf{H}\,;$$

therefore

$$\dot{\mathbf{q}} = 2\mathrm{A}c\mathbf{H}, \quad \text{curl } \mathbf{q} = 2\mathrm{A}\mathbf{D}\,;$$

so that the electrical energy is identified with the potential energy, or energy of ether twist, and the magnetic energy with the kinetic energy, or energy of ether flow, as stated on page 161.

Now let us write

$$\mathbf{u}=a\mathbf{H}$$

where a is a constant.

Then we have

$$W=\tfrac{1}{2}k\mathbf{u}^2=\tfrac{1}{2}ka^2\mathbf{H}^2, \quad T=\tfrac{1}{2}\rho\dot{\mathbf{q}}^2=2\rho a^2c^2\mathbf{D}^2.$$

For

$$a\dot{\mathbf{H}}=\dot{\mathbf{u}}=\tfrac{1}{2}\operatorname{curl}\dot{\mathbf{q}}, \quad \text{and} \quad a\dot{\mathbf{H}}=-ac\operatorname{curl}\mathbf{D};$$

therefore

$$\dot{\mathbf{q}}=-2ac\mathbf{D}, \quad \operatorname{curl}\mathbf{q}=2a\mathbf{H},$$

so that the electrical energy is now identified with the kinetic energy, or energy of ether flow, and the magnetic energy with the potential energy, or energy of ether twist.

It will be observed that in both these alternative representations we have

$$\operatorname{div}\dot{\mathbf{q}}=0,$$

which expresses the condition that the ether is incompressible; and

$$\ddot{\mathbf{q}}=-c^2\operatorname{curl}\operatorname{curl}\mathbf{q},$$

which is the known equation of motion of the M'Cullagh ether.

We now have to develop the consequences of each of these two alternative mechanical representations of the actions occurring in the electric field, in order to justify our selection of the former in preference to the latter.

Without confining ourselves to any special representation of the nature of an electron, such, for example, as the Larmor electron, which has been adopted in the text whenever it was advisable to have a definite physical concept of the electron presented to the reader's mind, we may, without any loss of generality, regard it as bounded by a surface of discontinuity in the otherwise continuous isotropic ether, regarded for the present, in order to simplify our analysis, as absolutely incompressible.

The problem of the transmission of the electron through the ether is then reduced to the hydrodynamical problem, which has been dealt with by Riemann, Hugoniot, and others, of the propagation of waves, or pulses, of discontinuity through an incompressible fluid medium. These investigations have ceased to be of physical interest, as far as material fluids are concerned, since Lord Rayleigh showed [1] that the existence of such waves would be inconsistent with the conservation of energy in any medium

[1] *Treatise on Sound*, Art. 253.

except one in which the adiabatics are straight lines, a condition which is not fulfilled in the case of any material fluid. It has been shown, however, by Rankine[1] that the straightness of the adiabatics is the necessary and sufficient condition for the permanency of type of continuous waves propagated in a fluid medium. Now this permanency of type is known, as was pointed out on page 419, to be characteristic of ether waves, and it therefore follows that the adiabatics of the ether must be straight lines, making it possible for such pulses of discontinuity to be transmitted through it. The relations between the ether displacements at the surface of discontinuity and the speed of propagation are determined, partly by the kinematical conditions which must be fulfilled in order to ensure stability and permanence of type, and partly by dynamical considerations depending on the nature of the medium and the structure of the electron.

I shall for the present follow Dr G. A. Schott[2] in applying the method developed by Hugoniot[3] for determining the kinematical conditions.

Let x, y, z be the co-ordinates of an ether particle at time t reckoned from an arbitrary initial instant at which the co-ordinates are a, b, c. Let $\partial/\partial t$ represent differentiation in the system x, y, z, t; and d/dt in the system a, b, c, t. Let u, v, w be the velocity of the particle at the time t; and let D be the Jacobian $d(x, y, z)/d(a, b, c)$, and D_{11}, etc., its minors. Then

$$\frac{d}{dt}=\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+w\frac{\partial}{\partial z}, \qquad \frac{\partial}{\partial x}=\frac{D_{11}}{D}\frac{d}{da}+\frac{D_{12}}{D}\frac{d}{db}+\frac{D_{13}}{D}\frac{d}{dc} \text{ etc.}$$

Let ϕ be a function of a, b, c, t, which, together with its differential coefficients up to the order $n-1$ inclusive, is continuous at the surface $f(a, b, c, t)=0$, while the differential coefficients of order n are discontinuous at this surface. Then if α, β, γ be the direction cosines of the normal to the surface drawn from 1 to 2, U the numerical value of the velocity $\mathbf{U}$ of a point on the surface along this normal, in the corresponding direction, and λ a factor independent of p, q, r, h; then Hugoniot's fundamental theorem is that, where $p+q+r+h=n$, and $[\phi']$ is written for brevity in place of $\phi'_2-\phi'_1$, where ϕ' is a function

[1] *Phil. Trans.*, vol. clx., 1870, p. 277; *Miscellaneous Scientific Papers*, p. 530; or Lamb's *Hydrodynamics*, p. 464.

[2] *Physikalische Zeitschrift*, vol. viii., 1907, p. 433. Dr Schott has kindly supplied me with portions of his argument, of which only the conclusions appear in this paper.

[3] *Journal de l'Ecole Polytechnique*, Cahier, 57, 1887, p. 1; *Jour. de Math.*, vol. iii., 1887, p. 477; or Hadamard, *Propagation des Ondes*, pp. 81–128.

which is discontinuous at the surface, and ϕ'_1 and ϕ'_2 are the values of ϕ' on opposite sides of the surface of discontinuity,

$$\left[\frac{d^n\phi}{da^p db^q dc^r dt^h}\right] = \lambda\alpha^p\beta^q\gamma^r(-\mathrm{U})^h.$$

Since λ, α, β, γ, U are, in general, functions of the co-ordinates and of the time, this shows that, in general, at all times, and at all points on the surface f, the differential coefficients of the nth order will experience a finite discontinuity, and the function ϕ may then be said to have a discontinuity of order n at the surface.

The assumption that the functions with which we have to deal are of the type ϕ, continuous functions of the co-ordinates, which enables us to apply Hugoniot's theorem to them, clearly involves the assumption that there is neither cavitation nor slip in the medium. It therefore confines the ether particles to the small displacements involved in electromagnetic action, and does not admit of their migration through the ether like the particles in a material fluid. This is in accordance with Lorentz's assumption of an ether without bodily motion, and which has no mechanical interaction with the electrons. We shall see presently that this assumption is untenable, but its provisional employment will be found to lead to conclusions which will indicate the nature of the requisite modification.

Since $\mathbf{q}$ is the ether displacement of a point P, arising from the twist $\mathbf{u}$, it is therefore the vector joining the point a, b, c, to the point x, y, z. We will suppose this displacement to be so small that higher powers than the first of $\mathbf{q}$ and its differential coefficients may be neglected. Let x, y, z be separated from a, b, c by a surface $f(a, b, c, t) = 0$, at which $\mathbf{q}$ has a discontinuity of order n.

Then, by Hugoniot's theorem, where $p+q+r+h=n$, and $\mathbf{Q}$ is a vector function of a, b, c, t, we shall have

$$\left[\frac{d^n\mathbf{q}}{da^p db^q dc^r dt^h}\right] = \mathbf{Q}\alpha^p\beta^q\gamma^r(-\mathrm{U})^h \quad . \quad . \quad (1).$$

From this it follows that, where $p+q+r+h=n-1$, and ν is the normal distance, to the first order, from the point to the surface, taken to be positive in the region 2,[1] and negative in the region 1, we shall have

$$\left[\frac{d^{n-1}\,\mathrm{curl}\,\mathbf{q}}{da^p db^q dc^r dt^h}\right] = \mathrm{V}\,.\,\nu\mathbf{Q}\alpha^p\beta^q\gamma^r(-\mathrm{U})^h \quad . \quad . \quad (2),$$

where $\mathrm{V}\,.\,\nu\mathbf{Q}$ represents, in the usual notation, the vector product of the vectors ν and $\mathbf{Q}$.

[1] See Hadamard, *Propagation des Ondes*, p. 103.

Schott points out further that, when Hugoniot's theorem is applied to the equations of Lorentz's electron theory,[1] it is found that continuity will be maintained if we make $\mathbf{U}=\mathbf{v}$, where $\mathbf{v}$ is the velocity of the electricity on the surface of the electron; which simply means that the surfaces of discontinuity for the potential and electric force must always coincide with the surface of the electron. It is also found that the vectors

$$\mathbf{d}'=\mathbf{D}+\frac{1}{c}\mathrm{V}.\mathbf{v}\mathbf{H}, \quad \mathbf{h}'=\mathbf{H}-\frac{1}{c}\mathrm{V}.\mathbf{v}\mathbf{D},$$

will have discontinuities only of order higher than those of $\mathbf{D}$ and $\mathbf{H}$. These results justify, as far as pure electromagnetic theory is concerned, Lorentz's attribution of free mobility to the electron, but it still remains to be determined whether this mobility is consistent with a mechanical representation of electromagnetic forces, which is essential if the electron theory is to be considered as a development of the general Faraday-Maxwell theory.

Taking, in the first place, $\mathbf{u}=\mathrm{A}\mathbf{D}$, that is to say, identifying the electrical energy with the potential energy of ether twist, and consequently, as we have seen, the magnetic energy with the kinetic energy of ether flow, we obtain as solutions of the fundamental equations in free ether

$$\mathbf{H}=\frac{1}{2\mathrm{A}c}\dot{\mathbf{q}}, \quad \mathbf{D}=\frac{1}{2\mathrm{A}}\operatorname{curl}\mathbf{q}.$$

Boltzmann,[2] assuming with Lorentz that the ether is not set in motion by the motion of the electron, has made the following objection to this representation:—Let a surface S be described so as to enclose completely an electron of charge e, while itself lying entirely in free ether, and let N be the outwardly directed normal component of the electric force at a point on this surface; then by Gauss's theorem

$$\iint \mathrm{N}d\mathrm{S}=e$$

when the integration is extended completely over the surface, a result which is inconsistent with the value ascribed to $\mathbf{D}$, according to which the integral should vanish. It will therefore be necessary to describe a small closed contour on the surface S, enclosing an area S′, and to confine the above integration to the surface S − S′, the integral over S′ having the value $-e$. If the distance of the surface S from the electron be increased, the area

[1] *Encyclopädie der math. Wiss.*, vol. V 2, pp. 147–8.
[2] *Ann. Phys. Chem.*, xlviii. p. 78, 1893.

S′ will trace out a narrow tubular space bounded by a tubular surface of discontinuity, the electric force within this surface being directed towards the electron, and away from the electron throughout external space. There must therefore be a tangential discontinuity in the electric force, and consequently in the ether twist measured by the curl of $\mathbf{q}$. The surface of discontinuity must therefore be the seat of tangential forces acting between the portions of ether on either side of it. On the other hand, there can be no corresponding permanent ether displacement, as if so there would be a resulting expenditure or production of energy. There can therefore be neither normal nor tangential discontinuity in the velocity, nor consequently in the vector $\mathbf{q}$.

Applying Hugoniot's theorem, we find from equations (1) and (2) that

$$[\dot{\mathbf{q}}] = -\mathrm{U}\mathbf{Q} \quad . \ . \ . \ . \ (3); \quad [\mathrm{curl}\,\mathbf{q}] = \mathrm{V}\,.\,\nu\mathbf{Q} \quad . \ . \ . \ . \ (4).$$

Let $\mathbf{Q}_\tau$ be the value of $\mathbf{Q}$ for a certain tangential direction τ,[1] then it follows from (3) and (4) respectively that

$$[\dot{\mathbf{q}}_\tau] = -\mathrm{U}\mathbf{Q}_\tau, \quad \text{and } \mathbf{Q}_\tau \neq 0.$$

But

$$[\dot{\mathbf{q}}_\tau] = 0,$$

and therefore

$$\mathrm{U} = 0.$$

That is to say, the normal velocity of the surface of discontinuity relatively to the ether must be zero everywhere. The electron forming the origin of the tubular surface can therefore only move in the direction of the axis of the tube. Now, the ether being incompressible, the corresponding lengthening or shortening of the tube can only be effected by setting the ether in motion. Now, according to the mechanical representation now being considered, $\mathbf{q}$ can only be decomposed into a sum of terms of the type $\dfrac{d^n\mathbf{q}}{da^p db^q dc^r dt^h}$, each of which will give $\mathrm{U}=0$, so that no terms in which h differs from zero can give rise to discontinuity in the electric force. It follows therefore that, on the assumption hitherto made that the properties of the ether inside and outside the electron are identical, the motion of an electron through the ether must set the latter in motion in its immediate neighbourhood, in very much the same manner that a material fluid is set in motion by the passage of a material body through it.

We will now make $\mathbf{u} = a\mathbf{H}$, thereby identifying the magnetic

[1] This direction must not be perpendicular to the discontinuity of $\mathbf{D}$, as in that case $\mathbf{Q}_\tau$ would vanish.

energy with the potential energy of ether twist, and consequently, as we have seen, the electrical energy with the kinetic energy of ether flow. We then obtain as solutions of the fundamental equations in free ether

$$\mathbf{D} = -\frac{1}{2ac}\dot{\mathbf{q}}, \quad \mathbf{H} = \frac{1}{2a}\operatorname{curl}\mathbf{q}.$$

We must again suppose a similar tubular surface to be described, and from the first of these two equations it must be a tube of ether flow, so that the tubular surface will be a surface of slip, and therefore the discontinuity of $\mathbf{q}$ will be of the first order and tangential, so that equations (3) and (4) will again hold good.

Let λ indicate a direction parallel to the axis of the tube and in the direction of the external ether flow, ν the outward normal, and μ a third direction at right angles to both, and therefore having a fixed direction along the circumference of the tube; then

$$[\dot{\mathbf{q}}] = -\mathrm{U}\mathbf{Q}_\lambda, \quad \mathbf{Q}_\mu = 0, \quad \mathbf{Q}_\nu = 0;$$

and therefore, from the first of the fundamental equations,

$$[\mathbf{D}] = -\frac{1}{2ac}[\dot{\mathbf{q}}] = \frac{1}{2ac}\mathrm{U}\mathbf{Q}_\lambda.$$

There must, therefore, be a tangential discontinuity along the surface of the tube, and this, according to Maxwell's equations of electric force, must give rise to surface electric currents; that is to say, on the basis of the representation now under consideration, to currents of bodily displacement in free ether.

Both these mechanical representations, therefore, necessitate the rejection of Lorentz's assumption that the ether, assumed to be similar within and without the electrons, is unaffected by the motion of the electrons through it, and this conclusion is supported by numerous other arguments.

In the first place, Lorentz's motionless ether, unaffected by the movements of electrons, would afford no basis for the Faraday-Maxwell mechanical theory; and, as Schott points out, this latter theory accounts far more satisfactorily for the observed results of Ampère, Faraday, Hertz, and others, than do Lorentz's equations, unless these are supplemented by mechanical theories inconsistent with his primary assumption.

Moreover, this assumption would preclude the only hypothesis yet suggested to account for the definite sizes of the electrons referred to on page 434 and discussed in Appendix P, and Poincaré[1] has pointed out that Lorentz's own proof of the

[1] *Archives Néerlandaises*, vol. v., 1900, p. 232.

principle of relativity makes the size of the electron undetermined, and renders it necessary, in accordance with the principle of the conservation of energy, to assume the existence of a mechanical pressure of the ether on the electron. He has also shown[1] that the fundamental dynamical principle of the equality of action and reaction would fail to hold, on the electromagnetic theory, unless the ether were the seat of kinetic energy, which could not very well be accounted for otherwise than in mechanical terms. Finally, the acceptance of Lorentz's assumption would preclude our attempting to account for gravitation by the hypothesis of ether flow, either continuous or alternating, which is considered in Chapter XXIII, and, as there explained, is the only one which appears to offer the possibility of accounting for gravitation.

Schott points out, moreover, that the only possibility of reconciling the mobility of the electron with a motionless ether would be by seeking for some modification of the ether within the electron as compared with the external ether which would be capable of invalidating the proofs which have been given of the incompatibility of a mobile electron with an immobile ether. Now, no such modification would affect Boltzmann's objection to the first representation. Again, the only modification which would obviate the difficulty in the case of the second representation would be to make the interior of the electron a void space, free from ether, making it therefore a vacuole in the ether; and there does not appear to be any possibility of accounting for the motion of such a vacuole through an immobile and incompressible ether.

Schott then proceeds to consider how the interaction between the electron and the surrounding ether may best be represented. The mechanical properties of the ether will be completely expressible in terms of its density and elasticity, and must be regarded as differing within and without the electron. The energy per unit volume of the external ether being $\frac{1}{2}(\mathbf{D}^2+\mathbf{H}^2)$, let that of the internal ether be $\frac{1}{2}(\mathrm{K}\mathbf{D}^2+\mu\mathbf{H}^2)$, where K and μ are, in general, functions of a, b, c, t.

Then, where ρ' is the volume density of electricity, the fundamental equations must be replaced by

$$\left.\begin{aligned} \rho'\frac{\mathbf{v}}{c}+\frac{1}{c}\frac{d}{dt}(\mathrm{K}\mathbf{D}) &= \operatorname{curl}\mathbf{H}, \; (a) \\ \frac{1}{c}\frac{d}{dt}(\mu\mathbf{H}) &= -\operatorname{curl}\mathbf{D}, \; (b) \\ \operatorname{div}(\mathrm{K}\mathbf{D})=\rho', \; (c); \; & \operatorname{div}(\mu\mathbf{H})=0, \; (d) \end{aligned}\right\} \quad . \quad . \quad (5)$$

[1] *Comptes Rendus*, 1905, p. 1504.

together with the equation expressing the invariability, throughout the motion, of the electric distribution, viz.,

$$\dot{\rho}' + \operatorname{div}(\rho' \mathbf{v}) = 0 .$$

We shall still assume, for the present, that neither slip nor cavitation occurs in the ether, so that Hugoniot's theorem will still be applicable.

If we write, for brevity, $K\mathbf{D} = \mathbf{p}$, (5*c*) becomes $\operatorname{div} \mathbf{p} = \rho'$, and applying it to two neighbouring points on opposite sides of the boundary and taking their differences, we have

$$(\operatorname{div} \mathbf{p})_2 - (\operatorname{div} \mathbf{p})_1 = \rho'_2 - \rho'_1 ,$$

or

$$[\operatorname{div} \mathbf{p}] = [\rho'].$$

Now ρ'_2 and ρ'_1, though different, are by assumption finite, and therefore $\operatorname{div} \mathbf{p}$ is finite.

Then, by Hugoniot's theorem, where $\mathbf{P}$ is some vector function of position, that is to say, some function of a, b, c, t, along the boundary,

$$\left[\frac{d\mathbf{p}}{da}\right] = \mathbf{P}\alpha , \quad \left[\frac{d\mathbf{p}}{db}\right] = \mathbf{P}\beta , \quad \left[\frac{d\mathbf{p}}{dc}\right] = \mathbf{P}\gamma , \quad \left[\frac{d\mathbf{p}}{dt}\right] = -\mathbf{P}\mathbf{v}_\nu ,$$

where α, β, γ are the direction cosines of the normal ν, and $\mathbf{v}_\nu$ is the normal velocity of the boundary.

It follows that

$$[\operatorname{div} \mathbf{p}] = \alpha \mathbf{P}_x + \beta \mathbf{P}_y + \gamma \mathbf{P}_z = \mathbf{P}_\nu ,$$

$$[\operatorname{curl} \mathbf{p}] = \beta \mathbf{P}_z - \gamma \mathbf{P}_y , \text{ etc.} = \mathrm{V} . \nu \mathbf{P} ,$$

where ν is the normal unit vector, so that (5*c*) becomes

$$\mathbf{P}_\nu = [\rho'] ,$$

so that $\mathbf{P}_\nu$ is finite everywhere along the boundary, and hence $\mathbf{P}_x$, $\mathbf{P}_y$, $\mathbf{P}_z$ are also finite, except possibly at certain singular points.[1]

Now the finiteness of $\mathbf{P}_x$, $\mathbf{P}_y$, $\mathbf{P}_z$ involves the finiteness of all such differences as

$$\left[\frac{d\mathbf{p}_x}{da}\right], \quad \left[\frac{d\mathbf{p}_x}{db}\right], \quad \left[\frac{d\mathbf{p}_x}{dc}\right], \quad \left[\frac{d\mathbf{p}_x}{dt}\right], \text{ etc.}$$

It follows, therefore, that for any tangential direction s, $[d\mathbf{p}/ds]$ is finite. Now the finiteness of a difference $[d\mathbf{p}/da]$ means either that both $(d\mathbf{p}/da)_1$ and $(d\mathbf{p}/da)_2$ are finite, or that both are infinite, but their difference finite. The latter alternative, except

[1] For example, $\mathbf{P}_x$ might be infinite for $\alpha = 0$, provided $\mathbf{P}_x\alpha$ remained finite, and so on; but such cases would be exceptional, and could not occur for the whole boundary, that is to say, for all values of α, β, γ.

at isolated singular points, appears to be excluded by physical considerations, as it does not appear possible to interpret physically the existence of such conditions at every point of the boundary. It therefore appears to be a legitimate conclusion that all the differential coefficients of the first order are finite on each side of the boundary, with the possible exception of isolated singular points.

Now, bearing in mind that separation and slip of the medium are excluded by assumption, $\mathbf{v}$ must be continuous at the boundary, and hence we have from (5*a*)

$$[\rho']\frac{\mathbf{v}}{c}+\frac{1}{c}\left[\frac{d\mathbf{p}}{dt}\right]=[\text{curl}\,\mathbf{H}].$$

Here the first term is necessarily finite, and $[d\mathbf{p}/dt]$ has been shown to be finite, so that $[\text{curl}\,\mathbf{H}]$ must also be finite. Again, we have as before, by Hugoniot's theorem

$$\left[\frac{d\mathbf{H}}{da}\right]=\mathbf{h}\alpha,\quad \left[\frac{d\mathbf{H}}{d\beta}\right]=\mathbf{h}\beta,\quad \left[\frac{d\mathbf{H}}{d\gamma}\right]=\mathbf{h}\gamma,\quad \left[\frac{d\mathbf{H}}{dt}\right]=-\mathbf{h}\mathbf{v}_\nu,$$

and

$$[\text{curl}\,\mathbf{H}]=\mathrm{V}\,.\,\nu\mathbf{h},$$

where $\mathbf{h}$ is some vector function of position along the boundary.

Hence we have

$$\mathrm{V}\,.\,\nu\mathbf{h}=[\rho']\frac{\mathbf{v}}{c}+\frac{1}{c}\left[\frac{d\mathbf{p}}{dt}\right]=[\rho']\frac{\mathbf{v}}{c}-\frac{\mathbf{v}_\nu}{c}\mathbf{P}.$$

Applying this to the normal direction we obtain

$$0=\left\{[\rho']-\mathbf{P}_\nu\right\}\frac{\mathbf{v}_\nu}{c},$$

which is simply (5*c*), showing that the equations are consistent, while for any two perpendicular tangential directions s and σ we have

$$\mathbf{h}_\sigma=[\rho']\frac{\mathbf{v}_s}{c}-\frac{\mathbf{v}_\nu}{c}\mathbf{P}_s,\qquad -\mathbf{h}\;=[\rho']\frac{\mathbf{v}_\sigma}{c}-\frac{\mathbf{v}_\nu}{c}\mathbf{P}_\sigma.$$

So far $\mathbf{h}_\nu$ and therefore $[d\mathbf{H}_\nu/da]$, etc., might be infinite; but by Hugoniot's theorem, if $\mathbf{h}_\nu$ be infinite, then every difference $[d\mathbf{H}_\nu/ds]$ must be infinite, for all directions of s. Let it be tangential; then this difference would have to be infinite on one or both sides of the boundary; that is to say, the normal magnetic force would be discontinuous, not merely in crossing the boundary, but also along it, which is impossible, and therefore $\mathbf{h}_\nu$ must be finite. It then follows as before that each first differential coefficient of $\mathbf{H}_x$, $\mathbf{H}_y$, $\mathbf{H}_z$ must be finite on each side of the boundary.

Remembering that $\mathbf{B}=\mu\mathbf{H}$, we find similarly from ($5d$) that all the first differential coefficients of $\mathbf{B}_x$, $\mathbf{B}_y$, $\mathbf{B}_z$ are finite; and therefore, from ($5b$), that all the first differential coefficients of $\mathbf{D}_x$, $\mathbf{D}_y$, $\mathbf{D}_z$ are finite.

Thus all the first differential coefficients of electric force and electric induction, and also of magnetic force and magnetic induction, are finite;[1] and since $\mathbf{p}=\mathrm{K}\mathbf{D}$, $\mathbf{B}=\mu\mathbf{H}$, this holds good also for the first differential coefficients of K and μ. Hence, in crossing the boundary, K, μ, $\mathbf{D}$, $\mathbf{p}$, $\mathbf{H}$, $\mathbf{B}$ are all continuous, because their first differential coefficients are finite, although the latter are discontinuous. At the boundary, therefore, except at singular points,

$$\mathrm{K}_1=\mathrm{K}_2=1\,,\quad \mathbf{D}_1=\mathbf{D}_2\,,\quad \mathbf{p}_1=\mathbf{p}_2\,,$$
$$\mu_1=\mu_2=1\,,\quad \mathbf{H}_1=\mathbf{H}_2\,,\quad \mathbf{B}_1=\mathbf{B}_2\,.$$

Now let

$$\left[\frac{d\mathrm{K}}{da}\right]=ka\,,\text{ etc.,}\quad \left[\frac{d\mu}{da}\right]=ma\,,\text{ etc.,}\quad \left[\frac{d\mathbf{D}}{da}\right]=\mathbf{d}a\,,\text{ etc.}$$

Then we have

$$\begin{aligned}\left[\frac{d\mathbf{p}}{dt}\right]&=\left[\mathrm{K}\frac{d\mathbf{D}}{dt}\right]+\left[\mathbf{D}\frac{d\mathrm{K}}{dt}\right]\\ &=\left[\frac{d\mathbf{D}}{dt}\right]+\mathbf{D}\left[\frac{d\mathrm{K}}{dt}\right]\\ &=-\mathbf{v}_\nu(\mathbf{d}+k\mathbf{D})\,,\end{aligned}$$

since at the boundary

$$\mathrm{K}_1=\mathrm{K}_2=1\,,\quad \mathbf{D}_1=\mathbf{D}_2=\mathbf{D}\text{ say.}$$

Hence ($5a$) becomes

$$[\rho']\frac{\mathbf{v}}{c}-\frac{\mathbf{v}_\nu}{c}(\mathbf{d}+k\mathbf{D})=\mathrm{V}\,.\,\nu\mathbf{h}\,.$$

The tangential equations are therefore

$$\mathbf{h}_\sigma=[\rho']\frac{\mathbf{v}_\varsigma}{c}-\frac{\mathbf{v}_\nu}{c}(\mathbf{d}_\varsigma+k\mathbf{D}_\varsigma)\,,\quad -\mathbf{h}_\varsigma=[\rho']\frac{\mathbf{v}_\sigma}{c}-\frac{\mathbf{v}_\nu}{c}(\mathbf{d}_\sigma+k\mathbf{D}_\sigma)\,;$$

which may be written

$$\mathbf{h}-\mathrm{V}\,.\,\frac{\mathbf{v}}{c}\mathbf{d}+[\rho']\mathrm{V}\,.\,\frac{\mathbf{v}}{c}\nu=k\mathrm{V}\,.\,\frac{\mathbf{v}}{c}\mathbf{D}\,.$$

[1] It is important to note that this result does not follow merely from the electromagnetic equations (5), but only as a consequence of the application to them of Hugoniot's theorem, on the assumption that there is neither separation nor slip in the medium.

Again, we have

$$\frac{d\mathbf{B}}{dt} = \left[\mu\frac{d\mathbf{H}}{dt}\right] + \left[\mathbf{H}\frac{d\mu}{dt}\right] = \left[\frac{d\mathbf{H}}{dt}\right] + \mathbf{H}\left[\frac{d\mu}{dt}\right] = -\mathbf{v}_\nu(h + m\mathbf{H}),$$

since at the boundary

$$\mu_1 = \mu_2 = 1, \quad \mathbf{H}_1 = \mathbf{H}_2 = \mathbf{H} \text{ say.}$$

Hence (5*b*) becomes

$$-\frac{\mathbf{v}_\nu}{c}(\mathbf{h} + m\mathbf{H}) = \mathrm{V}\,.\,\nu\mathbf{d}\,.$$

The tangential equations are therefore

$$\mathbf{d}_\sigma = -\frac{\mathbf{v}_\nu}{c}(\mathbf{h}_s + m\mathbf{H}_s), \quad -\mathbf{d}_s = -\frac{\mathbf{v}_\nu}{c}(\mathbf{h}_\sigma + m\mathbf{H}_\sigma);$$

which may be written

$$\mathbf{d} + \mathrm{V}\,.\,\frac{\mathbf{v}}{c}\mathbf{h} = -m\mathrm{V}\,.\,\frac{\mathbf{v}}{c}\mathbf{H}.$$

The electric and magnetic forces are not, in general, parallel to the velocity, so that the vector products $\mathrm{V}\,.\,\mathbf{vD}$ and $\mathrm{V}\,.\,\mathbf{vH}$ will not vanish, and therefore k and m will not become infinite. That is to say, there will be no discontinuity in K or μ, even in the case of a discontinuity in ρ' at the surface of the electron. It is, however, mechanically difficult to conceive how ρ', which is a function of K and μ, could become discontinuous at the surface of the electron while they remain continuous; or, generally, that ρ' could have, at the surface, a discontinuity of lower order than any discontinuity in K and μ.[1] If the interaction between the electron and the ether cannot be accounted for by a modification of the structure of the latter, it must be sought in a bodily displacement of the ether by the electron, resembling the displacement of material fluid by a material body passing through it.

We then have to admit slip in the medium, so that Hugoniot's theorem is no longer applicable. We have seen, however, that the introduction of slip, or irrotational lability,

[1] Dr Schott writes to the effect that, although difficult to imagine a mechanical explanation of such conditions, the demonstration can hardly be held to prove it absolutely inconceivable. In the last resort, ρ', K, and μ must all depend on the character of the ultimate motions to which the elasticity of the ether is due, and it appears quite conceivable that these might change with exceeding rapidity across a film of transition so narrow as to form an actual surface of discontinuity for the comparatively coarse-grained motions which give rise to electromagnetic phenomena. It is possible that K, μ might be continuous, while ρ' changed much more rapidly. He is not able, however, to see any way in which this could come about.

into the medium is all that is necessary to make the electron mobile through the ether according to the first of the two alternative representations considered. In the case of the second representation, on the other hand, we found that in a medium without slip there would be tangential discontinuities on the surface of the electron unless the nucleus consisted of a vacuole in the ether. It is not easy to see how these could be obviated by admitting slip, and their existence would appear necessarily to involve rotational motion, which is inadmissible. It is true that this difficulty would disappear if the nuclei consisted of vacuoles in the ether, as provisionally assumed by Sir Joseph Larmor. We found, however, that in a medium from which slip was excluded, this hypothesis was in itself inadmissible, independently of the mechanical representation adopted; and although the analysis by which this result was obtained ceases to be applicable when slip is admitted, we have no reason for concluding that the admission would remove the difficulty: the mobility of such an electron is therefore unproved.

The result to which our investigation leads us is that on the first representation, in which electric displacement is identified with ether twist, an electron in which the internal ether is similar to the external, or in which any difference between them is one which does not involve a discontinuity in ρ', is proved to be capable of unresisted motion through the ether. We are led to reject the second representation altogether, and the arguments considered on p. 187 against attributing magnetic effects to twistless ether flow are equally cogent against its giving rise to electric effects.

Dr C. V. Burton[1] contends that a self-locked strain-form cannot possess the property of free mobility through the ether, on the ground that circulation round a closed circuit, which is involved in its motion, must necessarily be accompanied by twist of the volume elements of the ether, which would give rise to resistance due to the elasticity of the ether. We have seen, however, in Chapter IX., that the high rotational elasticity of the ether entirely prevents the development of the twist which would be the necessary accompaniment of circulation in an incompressible perfect fluid without such rotational elasticity. The difficulty would present itself in the most acute form in the case of the continuous circulation involved, according to the mechanical representation adopted in the present work, in the existence of permanent magnets, and is, in my opinion, completely resolved by Sir Joseph Larmor's argument on p. 212.

[1] *Phil. Mag.*, vol. xiii., 1907, p. 705.

APPENDIX I.

THERMODYNAMICAL NOTES AND DEFINITIONS.[1]

Two experimental generalisations, known as the First, and Second, Law of Thermodynamics respectively, form the foundation of the subject.

The First Law is the expression of the observed facts that mechanical work can be transformed into heat, and that the ratio of transformation is found to be a fixed quantity. If H is the quantity of heat, measured in calories, which is obtained from a quantity W of mechanical work, measured in ergs, then the First Law may be expressed by the equation

$$W = JH,$$

where J is a definite number, known as the mechanical equivalent of heat. Its approximate value is $4 \cdot 19 \times 10^7$.

The Second Law may be expressed by the statement that *it is impossible, by inanimate material agency, to transform heat into mechanical work merely by cooling a body below the temperature of the coldest of surrounding objects.*[2] This law is the expression of the generally observed fact that, in every transformation of heat into mechanical work, the heat is found to be derived from a hotter body, and that, in the process of transformation, some of the heat so derived is invariably found to be transferred to a colder body, which is actually heated thereby.

Any arrangement for the transformation of heat into mechanical work is termed a *heat engine.*

The law expressing the conditions which determine the maximum possible *efficiency* of transformation of heat into mechanical work, that is to say, the maximum possible value of the ratio of

[1] A simple and concise outline of the fundamental principles of thermodynamics will be found in Poynting and Thomson's *Heat*, chapter xvii., and therefore these notes can be confined to brief reminders of the more important points.

[2] See Lord Kelvin, *Edinburgh Proc.*, 1851; *Math. Phys. Papers*, vol. i. p. 174.

the amount of heat taken from the hotter body, to the amount of mechanical work obtained, was shown by Sadi Carnot[1] to depend on a principle which he formulated in terms of the then accepted "caloric" theory of heat. This principle was first expressed in terms of the mechanical theory of heat by Clausius[2] in 1850, and, practically in the words employed above to express the second law of thermodynamics, by Lord Kelvin. The second law is therefore frequently referred to as *Carnot's Principle*, and sometimes as *Clausius's Principle*.

In order to avoid the use of the term temperature, the exact definition of which (p. 236) is based on the second law, the law may be stated in the form:—Heat cannot be transferred from a colder to a hotter body without the accompaniment of other correlative changes.

A reversible heat engine is one in which the result of reversing a complete cycle of operations—a complete cycle being one in which the bodies concerned are brought back to their original states—is the absorption of an amount of mechanical work equal to the amount expended in the direct cycle. Any process which can be represented as a sum of such reversible cycles is termed a reversible process.

If a body at temperature T receives or parts with an amount of heat energy Q, it is said to gain or lose entropy Q/T.

If the gain or loss of heat gives rise to a change of temperature, the process may be divided up into a number of steps throughout each of which the temperature may be regarded as practically constant, and the gain or loss of entropy will then be

$$\sum \frac{Q}{T} \quad \text{or} \quad \int \frac{dQ}{T}.$$

According to the definition of a reversible process, we must have in this case, as the quantitative expression of the second law,

$$\int \frac{dQ}{T} = 0 .$$

In all transformations of energy, whether reversible or irreversible, we also have, in accordance with the first law,

$$\int dQ = W .$$

[1] *Réflexions sur la puissance motrice du feu et sur les moyens propres à la développer*, Paris, 1824. On page 38 of this work he states that "La puissance motrice de la chaleur est indépendante des agens mis en œuvre pour la réaliser; sa quantité est fixée uniquement par les températures des corps, entre les quels se fait en dernier résultat le transport du calorique."

[2] *Ann. Phys. Chem.*, vol. lxxix., 1850, p. 500; *Phil. Mag.*, vol. ii., 1851, p. 102.

If the system considered can be completely specified by mechanical co-ordinates, as is the case for a homogeneous gas or other fluid, we have $dW = pdv$, where p, v are the pressure and the volume respectively, and therefore the expression of the first law becomes

$$\int dQ = \int pdv\,.$$

The work W expended or absorbed in a complete cycle can therefore be represented by the area of a closed curve of which p and v are the rectangular co-ordinates. Such a curve is known as an Indicator Diagram, as similar curves, drawn automatically by the action of mechanism, attached to steam- and gas-engine cylinders, are employed to indicate the work done in the cylinder during a complete cycle. A reversible engine must have a higher efficiency than that of an irreversible one working between the same temperature limits, for if the latter had the higher efficiency, the result of employing the former to work the latter in the reverse direction, so as to carry back heat from the colder body to the hotter one, would be to increase the heat in the latter at the expense of the former, so that the hot body would become hotter and the cold body colder without any expenditure of energy, which would contravene the second law. It follows from this that all reversible engines working between the same temperature limits must have the same efficiency. The efficiency of a reversible engine is therefore independent of the nature of the working substance. If, therefore, a quantity of heat Q_1 is taken from the hot body, or *source*, at temperature T_1, and a quantity Q_2 transferred to the cold body, or *refrigerator*, at temperature T_2, during a complete cycle of a reversible engine, the efficiency $(Q_1 - Q_2)/Q_1$, and therefore also Q_2/Q_1, will be a function of T_1, T_2, so that we may write

$$\frac{Q_2}{Q_1} = f(T_1,\ T_2)\,.$$

Now suppose the single reversible engine working between temperatures T_1 and T_2 to be replaced by two, working between the temperatures T_1 and T_3, and T_3 and T_2, respectively. Then we shall have

$$\frac{Q_3}{Q_1} = f(T_1,\ T_3), \quad \frac{Q_2}{Q_3} = f(T_3,\ T_2)\,;$$

and therefore, for all possible values of T_1, T_2, T_3,

$$f(T_1,\ T_2) = f(T_1,\ T_3)\,.\,f(T_3,\ T_2),$$

or

$$f(T_3,\ T_2) = \frac{f(T_1,\ T_2)}{f(T_1,\ T_3)}\,.$$

This ratio is therefore independent of T_1, and may be represented by $\phi(T_2)/\phi(T_3)$, so that

$$\frac{Q_1}{Q_2}=\frac{\phi(T_1)}{\phi(T_2)}, \quad \frac{Q_2}{Q_3}=\frac{\phi(T_2)}{\phi(T_3)}, \quad \frac{Q_3}{Q_1}=\frac{\phi(T_3)}{\phi(T_1)}.$$

The form of the function $\phi(T)$ will depend on the scale selected for the measurement of temperature, and we see from the above that this scale may be chosen so as to make $\phi(T)$ proportional to the temperature, in which case we shall have

$$Q_1 : Q_2 : Q_3 = T_1 : T_2 : T_3 .$$

T is then the absolute temperature as defined by Lord Kelvin,[1] viz. the absolute temperatures of any two bodies are in the same ratio as the quantities of heat lost or gained by the bodies when one acts as the source and the other as the refrigerator in a perfectly reversible cycle.

Since Q/T is the expression for the entropy, this result may be expressed in the form:—If a quantity of heat be allowed to descend a temperature slope, some of it being intercepted and transformed in its course, then, if the transformations are reversible, the entropy will remain constant throughout the slope.

Temperatures expressed on this absolute scale are found by observation[2] to vary approximately in the same proportion as the temperatures determined by the air thermometer, as might perhaps have been expected from consideration of the fact that the scale of a gas thermometer is very nearly independent of the particular gas used. The degrees may, therefore, be conveniently selected so that the interval between 0° and 100° C. may be represented by 100 degrees on either scale; and the volumes of a given quantity of gas at these two temperatures are very nearly in the ratio of 273 : 373, which is the meaning of the statement commonly made that the absolute zero of temperature is equal to − 273° C.

Any change in a body, which is unaccompanied by change in its temperature, is said to be *isothermal*.

A transference of heat is said to be *adiabatic* when, although there may be a gain or loss of heat, that is to say, of intrinsic energy, this is due solely to work done on or by the body, or bodies, and not to conduction or radiation.

[1] *Edinburgh Trans.*, vol. xxi., 1854, p. 125; *Math. Phys. Papers*, vol. i. p. 235.

[2] See Poynting and Thomson's *Heat*, p. 269.

APPENDIX J.

LARMOR'S THEORETICAL DEDUCTION OF CURIE'S LAW OF MAGNETISATION.

CONSIDER a mass of paramagnetic material, moved up from a place where the intensity of the magnetic field vanishes to where it is H. The aggregate per unit volume of the total magnetic energies of its molecules is thereby altered from null to $-\mathrm{IH}$ or $-\kappa\mathrm{H}^2$. The mechanical work done by the mass in virtue of its attraction by the field is $\frac{1}{2}\mathrm{IH}$, for the magnetisation is at each stage of its progress proportional to the magnetising force. Thus there remains a loss in the total magnetic energy of the molecules equal to $\frac{1}{2}\mathrm{IH}$; this can only have passed into heat in the material, for we work on the hypothesis that the field of force H is due to an absolutely permanent magnetic system, so that no energy is used up in producing magnetic displacements in the inducing magnets. Now let us apply Carnot's principle[1] to a reversible cycle in which the material is moved up into the field at temperature $\mathrm{T}+\delta\mathrm{T}$ and removed at temperature T, with adiabatic transition between these temperatures. Let $h+\delta h$ be the thermal energy per unit volume which it must receive from without at the higher temperature, and h that which it must return at the lower, in order to perform the amount of work

$$\delta\mathrm{W}=\tfrac{1}{2}\mathrm{H}^2\frac{d\kappa}{d\mathrm{T}}\,.\,\delta\mathrm{T}$$

in the cycle.

Then, by Carnot's principle,

$$\frac{\delta\mathrm{W}}{\delta\mathrm{T}}=\frac{h}{\mathrm{T}},$$

and, as above,

$$h=-\tfrac{1}{2}\kappa\mathrm{H}^2;$$

so that

$$\frac{d\kappa}{d\mathrm{T}}=-\frac{\kappa}{\mathrm{T}}.$$

Therefore, where A is a constant, $\kappa=\mathrm{A}/\mathrm{T}$, which is Curie's law.

[1] See Appendix I.

APPENDIX K.[1]

THE ETERNAL EXISTENCE OF AN UNCONTROLLED UNIVERSE IS INFINITELY IMPROBABLE.

It was mentioned on page 49 that the application of Lagrange's method to problems in which it is impossible to determine the co-ordinates completely will still afford all the information obtainable from the available data. All the problems of molecular physics are of this nature, for we are unable to determine the co-ordinates of the individual molecules, and therefore the process employed must necessarily involve the averaging of the effects due to the molecular motions, their averages being obtained from observations on matter in bulk. It is for this reason that problems in molecular physics are necessarily indeterminate, the existence of any one possible solution involving the possibility of an infinite number of others. The branch of abstract dynamics which is concerned with problems of this nature is usually known as *Statistical Mechanics*, and the method employed is as follows: All possible configurations of the system to be investigated are divided into mutually exclusive classes, $A_1, A_2, \ldots . A_n$. The corresponding classes of solution $B_1, B_2, \ldots . B_n$, after time t, corresponding to the initial configurations of the classes $A_1, A_2, \ldots . A_n$, are then calculated. The original classes are selected in such a manner that the values of any co-ordinate in any class A shall differ so slightly from one another that the final values of the same co-ordinate in the solution B shall also differ only slightly from one another. Suppose $p_1, p_2, \ldots . p_n$ to represent the unknown probabilities that the co-ordinates of the initial system belong to the classes $A_1, A_2, \ldots . A_n$, so that

$$p_1 + p_2 + \ldots . + p_n = 1.$$

[1] The substance of this Appendix, with the exception of the inference that the existence, as a conservative system, of a universe of ether and matter necessarily involves purposive directive intelligence, is contained in a paper by J. H. Jeans "On the Application of Statistical Mechanics to the General Dynamics of Matter and Ether" (*Proc. Roy. Soc.*, vol. lxxvi., A, 1905, p. 296).

Now assume that of the final classes $B_1, B_2, \ldots B_n$, $B_1, B_2, \ldots B_m$, where m is less than n, possess some special feature. Then the probability that the system will possess this feature after a time t is

$$p_1 + p_2 + \ldots + p_m,$$

and the probability that it will not possess this feature is

$$p_{m+1} + p_{m+2} + \ldots + p_n.$$

The odds in favour of the occurrence of this feature will then be represented by the fraction

$$\frac{p_1 + p_2 + \ldots + p_m}{p_{m+1} + p_{m+2} + \ldots + p_n} \quad . \quad . \quad . \quad (1)$$

Suppose now that N′ co-ordinates of the system all enter in exactly the same manner into the dynamical specification of the system, so that the energy function is symmetrical in these co-ordinates. Then the value of the fraction (1) will be a function of N′ and of certain constants $C_1, C_2, \ldots$ occurring in the specification of the system and in the probabilities $p_1, p_2, \ldots p_n$.

The procedure of statistical mechanics is to endeavour to find features for which the fraction (1) approaches the limit infinity when N′ is very great. When such a feature is found, it will be infinitely probable that it will be present in the system after the time t, and this infinite probability will be independent of the unknown law of probability of the initial co-ordinates. Jeans assumes that the systems considered are what Hertz calls *holonomic*.[1]

We may then assume, further, that the changes in the co-ordinates of the system are governed by equations of the canonical form

$$\dot{q}_r = \frac{\partial E}{\partial p_r}, \quad \dot{p}_r = -\frac{\partial E}{\partial q_r} \quad . \quad . \quad . \quad (2)$$

where q_r, p_r are a pair of corresponding co-ordinates of position and momentum, and E is the energy of the system, expressed as a function of the q's and $\dot{q}$'s.

The motion of the system may be most simply represented by a method which was, I believe, first employed by Professor W. K. Clifford, viz. supposing N′ to be the number of such pairs of co-ordinates, we imagine an ideal space of 2N′ dimensions, such that any 2N′ orthogonal axes in this space correspond to the values of the 2N′ independent co-ordinates

$$q_1, q_2, \ldots q_{N'}, \quad p_1, p_2, \ldots p_{N'}.$$

[1] See Appendix B.

Any configuration of the system can then be represented by a point in this space, the motion of the point being determined by the equations (2).

If the space is supposed to be filled with a continuous fluid, the different motions of the system, corresponding to all possible initial values of the co-ordinates, can be represented simultaneously. If the fluid is arranged so that, initially, masses $a_1, a_2, \ldots . a_n$ are found inside those parts of the space which represent the classes $A_1, A_2, \ldots . A_n$, and if the fluid move in accordance with equations (2), then, after the time t, the regions of the space representing the classes $B_1, B_2, \ldots . B_n$ will be occupied by the masses $a_1, a_2, \ldots . a_n$ of fluid. The probability that the system shall now possess the special feature considered will then be represented by the value of the fraction (1), viz.,

$$\frac{a_1 + a_2 + \ldots . + a_m}{a_{m+1} + a_{m+2} + \ldots . + a_n} \quad . \quad . \quad . \quad (3)$$

The fluid will necessarily move as though incompressible, for if ρ be its density, $D\rho/Dt$ the rate of increase of ρ as we follow an element of fluid in its motion, and u, v, w the velocity components of the element, then the ordinary three-dimensional equation of continuity will be,

$$\frac{D\rho}{Dt} + \rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) = 0\,,$$

and in the present $2N'$-dimensional case this becomes

$$\frac{D\rho}{Dt} + \rho\sum_{r=1}^{r=N'}\left(\frac{\partial \dot{p}_r}{\partial p_r} + \frac{\partial \dot{q}_r}{\partial q_r}\right) = 0\,,$$

and the right-hand member of this equation vanishes in consequence of equations (2). It follows that there is no accumulation of the fluid in any parts of the space, so that the volumes which represent the classes $B_1, B_2, \ldots . B_n$ are equal to those of the corresponding classes $A_1, A_2, \ldots . A_n$.

If the form of the function E, expressed as a function of the p's and q's, were known, the system would be completely determined, and it might be possible, by eliminating the time from equations (2), to obtain a number of integrals of the form

$$\left.\begin{aligned} \psi_1(p_1, p_2, \ldots . p_n, \quad q_1, q_2 \ldots . q_n) &= \text{constant} \\ \psi_2(p_1, p_2, \ldots . p_n, \quad q_1, q_2 \ldots . q_n) &= \text{constant, etc.} \end{aligned}\right\} \quad (4)$$

These integrals of the equations of motion represent families of surfaces in the $2N'$-dimensional space, and these surfaces have

the property that there can be no flow of fluid across them; they therefore act as water-tight compartments between which the fluid moves.

If there are no integrals of the form of equations (4), every function of $p_1, p_2, \ldots p_n, q_1, q_2, \ldots q_n$, will, in general, vary with the time, and there will be no limits to the possible extent of this variation. Now let χ be any function, for example, p_2, q_1, the energy, etc., such that those parts of the space in which χ is numerically less than a certain finite value χ_1, are finite in extent. Then, if a point be selected at random from the whole space, it is infinitely probable that the value of χ will be greater than χ_1.

Consider now a system, represented by a point in the ideal space, starting from any initial value of χ, say χ_0, and moving for an infinite time. Since nothing is known about the initial co-ordinates of the system, every point in the space will be a possible point for the representation of the ultimate state of the system; and, since there is no accumulation of the fluid in any region of the space, every point of the space has the same probability of representing the final state of the system. It follows that, for a system starting from an arbitrary initial value χ_0 of χ, it is infinitely probable that, after a sufficiently long time, the value of χ will be greater than any arbitrary value χ_1. That is to say, the tendency of any function possessing the properties ascribed to χ is to increase indefinitely in value. The special feature here is that $\chi - \chi_1$ shall be positive; and the fraction (3) becomes infinite through the number of terms in the denominator vanishing in comparison with the number in the numerator.

This tendency to indefinite increase, or *irreversibility*, as it is called, is a direct consequence of the absence of any tendency to aggregation on the part of the fluid in the ideal space, and this in turn is a consequence of the perfect reversibility of the original equations of motion. Such irreversible phenomena are therefore inseparable from statistical mechanics.

In our universe the only function known to possess this property of indefinite increase is the entropy, and, as far as we know, the other possible χ-functions do not show this tendency. The inference is that the assumption that there are no water-tight compartments in the ideal space in which our universe is represented by a point is incorrect. There must therefore be at least one integral of the form of equations (4). The existence of one such integral follows, in fact, from the form of equations (2); and if the function E involve the momenta through terms of the second degree only, this integral is the energy integral, and Jeans observes that it is consistent with our knowledge of the universe

to suppose that the integral exists in this form. The ideal space may therefore be supposed to be divided into water-tight compartments by the family of surfaces determined by the equation

$$E = \text{constant},$$

and these compartments may be subdivided to any extent by surfaces representing other integrals of the system

The energy E is necessarily positive, and therefore E must be infinite at infinity in the ideal space, the loci $E = \text{constant}$ being closed or tubular surfaces, which, in general, decrease in size as the value of E decreases. The moving representative point then no longer has access to the whole space, so that the function χ no longer shows a tendency to indefinite increase.

The phenomenon of irreversibility does not, however, disappear entirely, for, no matter how many integrals of the equations of motion there may be, it may still be true that, in a single water-tight compartment, some functions, χ, χ', have the same value, χ_1, χ_1', at all except an infinitesimally small fraction of all the points in the compartment, so that after the system has been in motion for an infinite time, it is infinitely probable that these functions will have the values χ_1, χ_1'. The tendency of the functions χ, χ' is therefore always to approach, and never to recede, from the values χ_1, χ_1', and this tendency represents an irreversible phenomenon. A good illustration of this irreversibility is afforded by a gas. The kinetic energy may then be treated as constant, and is equal, except for a constant multiplier, to the sum of the squares of the momenta.[1]

Now consider a function $f(p)$, determined by the condition that the number of momenta at any point of the ideal space, the values of which lie within the narrow limits p and $p+dp$, form a fraction $f(p)dp$ of the whole, and suppose this function to be tabulated for every point of the space. Then the function

$$\int_{-\infty}^{+\infty} f(p) \log f(p) dp,$$

which measures the value of $f(p)$ averaged for all the momenta at the point, is a definite function of the co-ordinates of the point, and vanishes at all except an infinitesimal fraction of the points of any compartment. It is, therefore, infinitely probable that its value will be zero in the final state of the gas, and as it can be shown that it can never have a negative value, it has an irreversible tendency to decrease. When it attains its ultimate zero value the

[1] See Jeans, *Dynamical Theory of Gases*, chapters iv. and v.

velocities of the gaseous molecules will be distributed according to Maxwell's law, a proposition which is utilised in Chapter XIV.

In a universe consisting of molecular matter, without ether, the total energy will be the sum of the kinetic energies of the molecules, and of the potential energy of the intermolecular forces; and here a similar irreversibility is found, and in the ultimate state of the system this energy will be distributed in such a manner that the average kinetic energy of each degree of freedom of every molecule is the same.[1]

In a universe consisting of ether only, without matter, the energy would be that of the trains of ether waves. These trains can differ only in wave-length and direction, so that, when these are known, each train will be specified dynamically by one pair of co-ordinates. If the universe is finite in extent, the boundaries must be perfect reflectors impervious to energy, and the trains of waves must then be replaced by the normal modes of vibration of the ether within the boundaries.[2] In either case, the motion can be completely determined when the initial values of the co-ordinates are known, so that the application of statistical methods is not strictly necessary. It is, however, instructive, as Jeans points out, to ascertain the nature of the information obtainable by their use. Since there is no interaction between the different wave-trains, or normal modes of vibration, the energy of each of these will remain constant, and if 2M′ is the total number of co-ordinates, there will be M′ equations expressing this constancy, and determining the compartments in the ideal space. These are not, however, the only integrals. The energy may be expressed in the form

$$2\mathrm{E}=\sum_{r=1}^{r=\mathrm{M}'}(a_r\dot{\phi}_r^2+b_r\phi_r^2),$$

the solutions of which are of the form

$$\phi_r=\mathrm{A}_r\sin(p_rt+\epsilon_r),\quad a_r\dot{\phi}_r=\mathrm{A}_r\sqrt{a_rb_r}\cos(p_rt+\epsilon_r),$$

where $p_r^2=b_r/a_r$, so that

$$\tan^{-1}\left(\frac{\phi_r\sqrt{a_rb_r}}{a_r\dot{\phi}_r}\right)=p_rt+\epsilon_r.$$

There are therefore M′ − 1 integrals of the form

$$\sqrt{\frac{a_r}{b_r}}\tan^{-1}\left(\frac{\phi_r}{a_r\dot{\phi}_r}\sqrt{a_rb_r}\right)-\sqrt{\frac{a_1}{b_1}}\tan^{-1}\left(\frac{\phi_1}{a_1\dot{\phi}_1}\sqrt{a_1b_1}\right)=\text{constant},$$

[1] See Jeans, *Dynamical Theory of Gases*, pp. 73 to 78.

[2] See Lamb's *Hydrodynamics*, chapter viii., or Thomson and Tait's *Natural Philosophy*, Arts. 337 *et seq.*

obtained by giving to r the values 2, 3, ... M′. There are, therefore, altogether 2M′ − 1 integrals, so that the compartments are reduced to lines.

A certain number of functions of the co-ordinates can be found which do not remain constant throughout the whole length of a line, but which have the same values at all except an infinitesimal fraction of the points in the space. For example, at all except an infinitesimal fraction of these points,

$$\sum_{r=1}^{r=M'} (a_r \dot{\phi}_r^2 - b_r \phi_r^2) = 0\,,$$

affording a proof by statistical mechanics of the well-known theorem that the electric energy tends to become everywhere equal to the magnetic energy, which is the electrical analogue of the equipartition theorem for gases.

Now consider a finite universe consisting of both matter and ether, and assume, to begin with, that the interaction between matter and ether is very slight, so that the co-ordinates may be regarded as approximately represented by the aggregate of the two sets of co-ordinates taken separately. Consider the series of functions

$$\psi_1,\ \psi_2,\ \ldots\ \chi_1,\ \chi_2,\ \ldots\,,$$

of which the ψ's and the χ's represent the functions of the co-ordinates of the material system and the ether, respectively, which would remain constant in the absence of the other system. Then the whole series, involving the presence of both matter and ether, will remain nearly, but not quite, constant. None of the ψ's can remain quite constant, for they are known to be the energy and the six components of momentum of the material system, and the existence of the radiation pressure (see Chapter XII.) shows that all these quantities are altered in value by ether waves. It is also very improbable that any of the χ's should remain quite constant, for this would involve the existence of modes of ether vibration which are unaffected by the presence of matter, which does not appear conceivable when we remember that the presence of matter diminishes the elasticity of the ether in its neighbourhood. Moreover, if such modes of vibration existed, they would be imperceptible, since occurrences in the ether can be perceived by us only through their influence on matter. They can, therefore, have no phenomenal existence.

There is one, and only one, quantity, the total energy of matter and ether, which is known to remain constant during the motion, and if there be any others, their constancy would not appear to

affect any of the conclusions to be arrived at. Let N′ and M′ be the number of degrees of freedom of the matter and the ether, respectively. The system will then have 2N′ co-ordinates which very nearly represent motion of matter only, and 2M′ co-ordinates which very nearly represent motion of ether only. If the ether be absolutely continuous, entirely without granular structure, M′ will be absolutely infinite, but in any case it will be very large in comparison with N′.

If we neglect the small terms arising from interaction, the energy of the ether may be expressed as the sum of 2M′ squares, and the energy of the matter may be divided into kinetic and potential energy. Assuming the number of electrons moving with velocities comparable with the velocity of light to be negligible in comparison with the total number, this kinetic energy may also be expressed as a sum of squares, viz. as a sum of the squares of the three components of momentum of each electron of which the matter is composed. If the electrons be sufficiently near to one another, the expression for the kinetic energy will be modified by the introduction of products of the components of momenta in pairs, but being still of homogeneous quadratic form, it can be immediately transformed into a sum of squares. The total energy is therefore expressible as a sum of 2M′ + N′ squares, together with an unknown potential energy of electrons. It will then follow, as in the general dynamical proof of Maxwell's law,[1] that throughout the compartment of the ideal space in which the universe must now be represented, the sum of any p of these squares bears to the remaining q squares a ratio which is equal to p/q at all except an infinitesimal fraction of the points in the compartment, subject only to the condition that p and q are large enough to be treated as infinite without appreciable error, conditions which are satisfied by 2M′ and N′. The system therefore tends to attain a state in which the energy of the ether is infinite compared with the kinetic energy of the matter, so that there is a general tendency for the ether to gain energy at the expense of the matter.

Our own universe is of the type last considered, consisting of ether and matter. It could not therefore, if uncontrolled, form a conservative system. It would necessarily have had an origin in time, and necessarily be approaching towards a state of physical death in which no further phenomenal change would be possible. This state of things could only be obviated by the application of external control to the changes in energy distribution, of such a nature as would compel the representative point in the ideal space

[1] Jeans, *Dynamical Theory of Gases*, chapter iii.

to follow a path which would be an infinitely improbable one in the absence of such control. It is inconceivable that such control should be exerted by any unintelligent, purposeless action, for it would be necessary for the moving representative point, except for possible divagations extending only to an infinitesimal fraction of the compartment of the ideal space, to move, at each instant of the duration of the universe, along one of a series of paths forming an infinitesimal fraction of the paths open to it in the absence of control. It is shown, moreover, in Chapter XXIV., that control of precisely similar character, although on a much smaller scale, is exercised by every living being known to us, the amount of such control increasing with the intelligence of the living agent. The exercise of such control is therefore ascertainable by ordinary observation.

We have seen that a similar argument would apply to a universe of matter without ether, if such a universe could exist. The existence of such a universe can, however, scarcely be regarded as conceivable.

APPENDIX L.

ELECTROMOTIVE FORCE OF DIFFUSION.

NERNST has shown that when steady diffusion is allowed to take place in an electrolyte traversed by an electric current, the back E.M.F. at the junction of the fluids is thereby reduced in the ratio of the difference of the ionic mobilities to their sum.

Let v_1 and v_2 represent what Kohlrausch calls the mobilities of the cathion and the anion respectively, that is to say, their mean velocities under unit electric force, and let n be the number of free positive ions, which is also the number of free negative ions, in unit of volume, e the charge on each ion. Consider now the diffusion across a layer in which the concentration varies, when the steady state is attained, so that both ions diffuse together at the same rate. The mean steady velocity of an ion will be equal to this mobility divided by its electric charge and multiplied by the force causing the motion. This force will consist of an electric part and an osmotic part. If V is the electric potential which has been set up during the transition to the steady state of diffusion, the electric part of the force will be $-e\dfrac{dV}{dx}$. Now, in order to maintain a steady state without diffusion by means of a porous partition, the partition must exert a greater pressure on the solution than on the pure solvent by an amount p. Considering a layer of the actual solution, of unit cross-section, and thickness δx, a bodily force $\dfrac{dp}{dx}\delta x$ would have to be exerted on it to prevent diffusion. $-\dfrac{dp}{dx}\delta x$ therefore represents the total forcive arising from the gradient of concentration, and acting on the ions to produce diffusion. The mean force per ion will therefore be $-\dfrac{1}{n}\dfrac{dp}{dx}$. Let $\dfrac{dN}{dt}$ be the number of single ions of either kind driven across unit area by these forces, then

$$\frac{dN}{dt} = -nv_1\frac{dV}{dx} - \frac{n}{e}\,\frac{v_1}{n}\,\frac{dp}{dx},$$

also

$$\frac{dN}{dt} = nv_2\frac{dV}{dx} - \frac{n}{e}\frac{v_2}{n}\frac{dp}{dx}.$$

Now, if D is the coefficient of diffusion of the solution,

$$\frac{dN}{dt} = -D\frac{dn}{dx}.$$

Moreover, if the solution is sufficiently dilute, $p = neRT$, where T is the absolute temperature and R is the constant of a perfect gas, and is independent of the nature of the ions. Therefore $\frac{dp}{dx} = eRT\frac{dn}{dx}$, and therefore the above equations give

$$D = \frac{2v_1v_2}{v_1+v_2}RT, \quad \frac{dV}{dx} = \frac{v_2-v_1}{v_2+v_1}\frac{1}{ne}\frac{dp}{dx};$$

so that the integrated potential difference $V_2 - V_1$ across the diffusion layer is

$$\frac{v_2-v_1}{v_2+v_1}RT\log\frac{p_2}{p_1},$$

which proves the statement.

APPENDIX M.

NOTES ON PROFESSOR J. H. JEANS' CONTRIBUTIONS TO THE ELECTRON THEORY OF METALLIC CONDUCTION AND RADIATION.

MAXWELL's law of the equipartition of energy between the different degrees of freedom of the molecules of a gas, which forms the basis of the treatment of the subject in Chapter XIV., has already been referred to in Appendix B, and has been very fully discussed by J. H. Jeans.[1] It determines a law of distribution of the velocities of the molecules of a gas about their mean value, which is, on the whole, unchanged by collisions, and is such that, with an initial random distribution of these velocities, they will tend, after a sufficient time, towards a distribution in accordance therewith. If m be the mass of a molecule and u, v, w its velocity components in the final state, and N the number of molecules per unit volume of a gas at a temperature T determined by the equation

$$\mathrm{RT}=\frac{1}{2h},$$

then the law of distribution of velocities is

$$\mathrm{N}\sqrt{\frac{h^3m^3}{\pi^3}}e^{-hm(u^2+v^2+w^2)}du\,dv\,dw, \quad . \quad . \quad . \quad (1)$$

where e is the base of the natural system of logarithms. That is to say, this expression represents the probability that the velocity of any one molecule selected at random shall have components lying between u and $u+du$, v and $v+dv$, w and $w+dw$ respectively. We now assume, as stated on p. 298, that this law determines the distribution of the free electrons in a conductor which is under the influence of no externally impressed electric force.

If an electric force X be made to act parallel to OX, this dis-

[1] *Dynamical Theory of Gases*, chapters ii. to v.

tribution will be changed. Each electron of charge e will now acquire momentum parallel to OX at a rate Xe, but this gain of momentum will be held in check by a continuous transfer of momentum between each electron and all the molecules by which it is influenced at any instant.

In the original state the average value of u is zero, but under the new conditions it will have a value u_0, differing from zero; and corresponding to any value of u_0 there will be a current C_x parallel to OX of amount Neu_0. It was shown on p. 302, by means of the assumption that the motion of an electron consists of free paths and collisions, that this current will obey Ohm's law for any electric forces within our experience.

Jeans[1] has shown that the same result may be obtained without assuming the existence of free paths. The calculations on pp. 303–4 show that, for all values of X within the limits of our experience, u_0 is small in comparison with the average numerical values of u. We may therefore neglect squares of u_0, and in calculating any quantity which has ultimately to be multiplied by u_0, we may assume a distribution as in (1).

An expression for the transfer of momentum between electrons and molecules is, for simplicity, first obtained on the assumption that the motion consists of free paths and collisions. An electron approaching a molecule with velocity components u, v, w will describe a curved orbit having its free paths before and after collision as asymptotes. The loss of momentum during the collision will depend on the velocity, the orientation of the molecule, and the point at which the free path before collision intersects a perpendicular plane through the molecule. All positions are equally likely for this point, and if the conductor is isotropic all orientations are equally likely for the molecule. The vector representing the average loss of momentum can therefore have no other direction than that of the initial velocity u, v, w, and therefore its amount will depend on the constants of the electrons and molecules and on $u^2+v^2+w^2$, but not on u, v, w, separately. We may therefore consider u, v, w as being reduced by the collision to

$$(1-a)u,\ (1-a)v,\ (1-a)w,$$

where a depends on $u^2+v^2+w^2$ and constants only.

The loss of momentum of the N electrons, in a time dt which is large compared with the time of a collision, is therefore of the form

$$N\gamma u_0 dt,\quad N\gamma v_0 dt,\quad N\gamma w_0 dt,$$

[1] *Phil. Mag.*, vol. xvii., 1909, p. 773.

where u_0, v_0, w_0 are average values of u, v, w, and γ depends on $u^2+v^2+w^2$ and constants only.

Therefore, provided the interval of time dt is taken to be large compared with the time of collision,

$$\frac{d}{dt}(\mathrm{N}mu_0)=\mathrm{N}\mathrm{X}e-\mathrm{N}\gamma u_0 \quad . \quad . \quad . \quad (2)$$

To effect the calculation without assuming the existence of free paths, consider the group of electrons in a conductor unacted on by external electric force which, at any given instant $t=0$, have velocity components lying within a small range du, dv, dw surrounding the values u, v, w. Their law of distribution in space will then be

$$\mathrm{NA}e^{-2h\chi}dxdydz, \quad . \quad . \quad . \quad . \quad (3)$$

where χ is the potential energy of an electron at the point x, y, z and A is determined by the equation

$$\mathrm{A}\iiint e^{-2h\chi}dxdydz=1\,,$$

the integral being taken throughout a unit volume. In the presence of an electric force X, the law will differ from this by terms of the order u_0/u, but these terms may be neglected whenever their retention would ultimately lead to terms of the order u_0^2/u^2. We may therefore assume the spatial distribution of these electrons to be in accordance with (3), and calculate accordingly the average values of u, v, w after a time τ corresponding to all initial positions of the electrons. Let these be u', v', w', then the vector u', v', w' must, from symmetry, be in the same direction as u, v, w, and its amount will depend on $u^2+v^2+w^2$ and the constants of the matter and electrons.

Hence

$$u'=uf(\tau,\,u^2+v^2+w^2),\ \text{etc.};$$

and therefore, on averaging for all values of u, v, w, we find that if u'_0, v'_0, w'_0 are the values of u_0, v_0, w_0 after the time τ, then

$$u'_0=u_0\phi(\tau,\,h) \quad . \quad . \quad . \quad . \quad (4)$$

This equation is a generalisation, in integral form, of (2) with $\mathrm{X}=0$.

Now, at time $t=0$

$$m\frac{du}{dt}=-\frac{d\chi}{dx},$$

so that, on averaging, the value of du_0/dt is zero, and hence, for small values of τ, $\phi(\tau, h)$, is of the form $1+\tau^2 f(h)$.

Again, suppose τ to be so large, compared with the time of collision with a molecule, that the velocities of those electrons which originally had velocities u, v, w may be regarded as distributed at random. Then, if u''_0, v''_0, w''_0 are the values of u_0, v_0, w_0 after a time 2τ, we shall have, in addition to (4),

$$u''_0 = u_0\phi(2\tau, h), \quad u''_0 = u'_0\phi(\tau, h), \text{ etc.,}$$

and therefore

$$\phi(\tau) = Ae^{-B\tau};$$

so that if dt be sufficiently large, (4) may be expressed in the form

$$\frac{du_0}{dt} = -Bu_0 .$$

This is exactly of the form of (2) with $X=0$, and therefore the general equation, no matter what the nature of the motion of the electrons, is of the same type as (2). Equation (2), with a suitable value for γ, will therefore express the relation between X and u_0 in any motion in which the quantities do not change too rapidly with the time. In the steady state in which a steady current C_x is maintained by an electric force X, the left-hand member must vanish, so that we have

$$u_0 = \frac{Xe}{\gamma},$$

and therefore, since $C_x = Neu_0$,

$$C_x = \frac{Ne^2}{\gamma}X,$$

which is in accordance with Ohm's law.

The conductivity κ for steady currents will therefore have the value Ne^2/γ, and therefore (2) may be written

$$\frac{dC_x}{dt} = \frac{Ne^2}{m}\left(X - \frac{C_x}{\kappa}\right) \quad . \quad . \quad . \quad . \quad (5)$$

If the steady electric force X be replaced by a periodic impressed electric force of period p by writing $X = X_0 \cos pt$, then the solution of this equation will be

$$C_x = \kappa X_0 \cos(pt - \epsilon)\cos\epsilon ,$$

where

$$\tan\epsilon = \kappa p \frac{m}{Ne^2}.$$

To calculate the rate at which energy is dissipated by the resistance, we note that in time dt and volume $dxdydz$ a quantity of electricity $C_x dtdydz$ falls through a P.D. $X_0 \cos ptdx$, its gain in momentum being nil. The energy dissipated is therefore

$$C_x X_0 \cos ptdxdydzdt,$$

so that the rate of dissipation per unit volume per unit time is

$$\begin{aligned} C_x X_0 \cos pt &= \kappa X_0^2 \cos \epsilon \cos (pt - \epsilon) \cos pt \\ &= \kappa X_0^2 (\cos^2 \epsilon \cos^2 pt + \cos \epsilon \sin \epsilon \cos pt \sin pt). \end{aligned}$$

The mean value of this averaged over a number of complete periods is

$$\tfrac{1}{2}\kappa X_0^2 \cos^2 \epsilon = \tfrac{1}{2} X_0^2 \frac{\kappa}{1 + \dfrac{\kappa^2 p^2 m^2}{N^2 e^2}}.$$

If k is the conductivity for currents of frequency p, this expression must be equal to $\tfrac{1}{2}kX_0^2$, so that we find, as stated on p. 300,

$$k = \frac{\kappa}{1 + \dfrac{\kappa^2 p^2 m^2}{N^2 e^2}}.$$

Now, where X is the x-component of electric force in electromagnetic units, C_x the corresponding component of electron current, related to X by (5), and c is the constant of radiation, then the total x-component of current is

$$\frac{K}{4\pi c} \frac{dX}{dt} + C_x,$$

so that this expression is equal, by Maxwell's equations, to

$$\tfrac{1}{4}\pi \left(\frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} \right),$$

where β and γ denote, as usual, components of the magnetic force **H**.

For electric waves of frequency p, X and C_x may both be taken as proportional to e^{ipt}, so that from (5)

$$C_x = \frac{1}{ip\left(\dfrac{1}{\kappa} + \dfrac{m}{Ne^2} ip \right)} \frac{dX}{dt},$$

and therefore

$$\frac{K}{4\pi c} \frac{dX}{dt} + C_x = \frac{K'}{4\pi c} \frac{dX}{dt},$$

where

$$K' = K + \frac{4\pi c}{ip\left(\frac{1}{\kappa} + \frac{m}{Ne^2}ip\right)} \quad . \quad . \quad . \quad . \quad (6)$$

The analysis for the propagation of light in conductors is therefore identical in form with that for the propagation in a non-conducting medium of dielectric constant K′, the equations of propagation being identical with those obtained by Drude in a different manner; and the present manner of their derivation

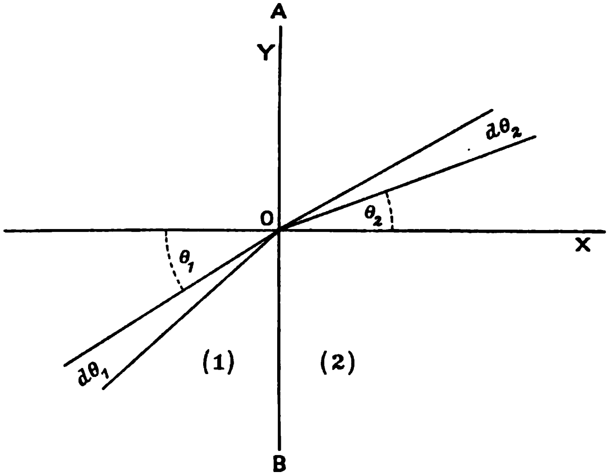


FIG. 34.

shows them to be applicable only to electric waves of periods much greater than the time of a collision between an electron and a molecule. Subject to this limitation, we have now proved the statement made on page 256, that "the solutions of problems involving reflection and refraction in absorbing media can be derived from the solutions obtained for transparent media by replacing the real dielectric constants by complex ones."

We are now in a position to prove Kirchhoff's law, which was stated on page 320. Let AB, fig. 34, be the boundary separating (1) the medium air from (2) a metallic medium, and suppose a beam of radiation of frequency p to be incident on AB, bounded by cones of angles θ_1 and $\theta_1 + d\theta_1$ in the first medium, and θ_2 and $\theta_2 + d\theta_2$ in the second.

Then, where K_1 and K_2 are the dielectric constants of the respective media, the equations of propagation will be

$$\frac{\partial\gamma}{\partial y}-\frac{\partial\beta}{\partial z}=\frac{K_1}{c}\frac{dX}{dt}, \text{ etc.,}$$

in air, and in the metallic medium

$$\frac{\partial\gamma}{\partial y}-\frac{\partial\beta}{\partial x}=\frac{K_2}{c}\cdot\frac{dX}{dt}+4\pi C_x=\frac{K'_2}{c}\cdot\frac{dX}{dt}, \text{ etc.,}$$

where the relation between K'_2 and K_2 is determined by (6).

The velocity v_1 of radiation in medium (1) will be determined by the equation $v_1^2=c^2/K_1\mu_1$. Now introduce an angle θ'_2 and a velocity v'_2 determined by the equations

$$\frac{\sin^2\theta'_2}{\sin^2\theta_1}=\frac{\dfrac{1}{K'_2\mu_2}}{\dfrac{1}{K_1\mu_1}}=\frac{v'^2_2}{v_1^2} \quad . \quad . \quad . \quad (7)$$

Then $\sin\theta'_2$ and v'_2 will both be complex quantities, but their ratio is real.

The exponential through which the time and space co-ordinates are expressed in the radiation in medium (1) may be taken to be

$$\exp.\ ip\left\{t-\frac{1}{v_1}(x\cos\theta_1+y\sin\theta_1)\right\},$$

and the corresponding expression for medium (2) will be

$$\exp.\ ip\left\{t-\frac{1}{v'_2}(x\cos\theta'_2+y\sin\theta'_2)\right\}.$$

Since $\sin\theta'_2/v'_2$ is real, it will be possible to introduce two new real quantities θ_2 and v_2 such that

$$\frac{\sin\theta'_2}{v'_2}=\frac{\sin\theta_2}{v_2}, \quad \frac{\cos\theta'_2}{v'_2}=\frac{\cos\theta_2}{v_2}-i,$$

the exponential for (2) now becomes

$$e^{-\iota px}\exp.\ ip\left\{t-\frac{1}{v_2}(x\cos\theta_2+y\sin\theta_2)\right\},$$

showing that θ_2 and v_2 are, respectively, the angle of refraction and velocity of radiation in medium (2).

Now introduce another complex quantity u_{12}, associated with

radiation passing from medium (1) to medium (2), and given, for vibrations polarised in the plane of incidence, by the equation

$$u_{12}{}^2 = \frac{K'_2}{\mu_2}\frac{\mu_1}{K_1}\frac{\cos^2\theta'_2}{\cos^2\theta_1},$$

and for vibrations polarised perpendicular to the plane of incidence, by the equation

$$u_{12}{}^2 = \frac{\mu_2}{K'_2}\frac{K_1}{\mu_1}\frac{\cos^2\theta'_2}{\cos^2\theta_1}.$$

Let R_{12} denote the coefficient of reflection of radiation incident at the angle θ_1 in medium (1), that is to say, the ratio of the amplitude of the reflected and incident rays. Then it can be shown in the usual way[1] that

$$R_{12} = \text{mod.}\ \frac{1-u_{12}}{1+u_{12}},$$

and similarly that for the coefficient of reflection of radiation incident at the angle θ_2 in medium (2)

$$R_{21} = \text{mod.}\ \frac{1-u_{21}}{1+u_{21}}$$

where u_{21} is defined by the equations corresponding to those which determine u_{12}. Then $u_{21} = 1/u_{12}$, and therefore $R_{12} = R_{21}$. It must be noted that while R_{12} is evaluated for light incident at the angle θ_1, R_{21} is evaluated for light incident at the angle θ_2.

Let both media be filled with radiation such that the energy per unit volume of radiation of frequency between p and $p+dp$ is E_1dp in (1) and E_2dp in (2). Then the stream of energy falling on the boundary per unit time at angles between θ_1 and $\theta_1+d\theta_1$ is

$$\tfrac{1}{2}E_1v_1\cos\theta_1\sin\theta_1d\theta_1dp,$$

so that the amount transmitted from (1) to (2) is

$$\tfrac{1}{2}E_1v_1dp\int(1-R_{12})\cos\theta_1\sin\theta_1d\theta_1 \quad . \quad . \quad (8)$$

Similarly, the amount transmitted from (2) to (1) is

$$\tfrac{1}{2}E_2v_2dp\int(1-R_{21})\cos\theta_2\sin\theta_2d\theta_2 \quad . \quad . \quad (9)$$

[1] Drude's *Lehrbuch der Optik*, p. 334 ; or Jeans' *Electricity and Magnetism*, p. 523.

In a steady state these two expressions must be equal. Now from (7)

$$\frac{\sin\theta_1}{v_1}=\frac{\sin\theta'_2}{v'_2}=\frac{\sin\theta_2}{v_2},$$

and therefore

$$\frac{\cos\theta_1\sin\theta_1 d\theta_1}{v_1^2}=\frac{\cos\theta_2\sin\theta_2 d\theta_2}{v_2^2}.$$

But since R_{12} is the same function of θ_1 as R_{21} is of θ_2,

$$\frac{1}{v_1^2}\int(1-R_{12})\cos\theta_1\sin\theta_1 d\theta_1=\frac{1}{v_2^2}\int(1-R_{21})\cos\theta_2\sin\theta_2 d\theta_2 \quad . \quad (10)$$

so that (8) and (9) will be equal if

$$E_1 v_1^3 = E_2 v_2^3.$$

The energy in a cavity in any kind of matter being of the form

$$E_1 = f(T\,.\,p),$$

it follows that the energy within the matter itself must be of the form

$$E_2=\left(\frac{v_1}{v_2}\right)^3 f(T,p), \quad . \quad . \quad . \quad (11)$$

so that the partition of energy inside matter can be determined in terms of black-body radiation. Substituting this value of E_2 in (9), we obtain as the total emission per unit area per unit time from the surface of the body

$$dp f(T,p)\frac{v_1^3}{2v_2^2}\int(1-R_{21})\cos\theta_2\sin\theta_2 d\theta_2,$$

and by (10) this is equal to

$$dp f(T,p)\tfrac{1}{2}v_1\int(1-R_{12})\cos\theta_1\sin\theta_1 d\theta_1.$$

Now the limits for θ_1 are from 0 to $\pi/2$, and therefore the coefficient of absorption A for radiation incident on the body is given by the equation

$$A=\frac{\int_0^{\frac{\pi}{2}}(1-R_{12})\cos\theta_1\sin\theta_1 d\theta_1}{\int_0^{\frac{\pi}{2}}\cos\theta_1\sin\theta_1 d\theta_1}$$

$$=2\int_0^{\frac{\pi}{2}}(1-R_{12})\cos\theta_1\sin\theta_1 d\theta_1,$$

and therefore the stream of issuing radiation is

$$\tfrac{1}{4}\mathrm{A}f(\mathrm{T}, p)dp.$$

Since $\mathrm{A}=1$ for a perfectly black body, it follows that the stream of radiation issuing from any body is A times the stream issuing from a black body, which is Kirchhoff's law.

We have now to prove the statement made on p. 321 that the transfer between matter and ether of the energy of high frequency vibrations is extremely slow. Jeans has shown[1] that when a vibration of any dynamical system is influenced by an external agency, the ultimate effect of this influence is infinitesimal, except when the external agency changes to a considerable extent in a time comparable with the period of the vibration. If the time of change in the external agency is n times the period of the vibration, where n is large, then the ultimate change in the energy of the vibration vanishes to the same order as e^{-n}, a quantity which soon becomes negligible as n increases. Suppose θ to be an interval of time so small that the material system may be regarded as perceptibly unaltered during such an interval, then the change produced in the energy of ether vibrations of periods less than θ will be extremely small, and therefore, if our consideration of the system does not extend over a very long period, the energy of such vibrations may be treated as though it were incapable of change.

It follows, as Jeans observes,[2] that although the total number of modes of vibration of any enclosed or unenclosed region of ether is either very great or infinite, as was pointed out in Appendix K, the number of vibrations of frequencies below an assigned value, of an enclosed region of ether, is finite. The quantity M′ of Appendix K may therefore be replaced by a smaller number M'_1, which will always be finite. If then we limit our consideration of the system to a finite time, the equations expressing the constancy of the energies of the remaining modes of vibration may be regarded as determining water-tight compartments in an ideal space of only $2(\mathrm{M}'+\mathrm{N}')$ dimensions. The ratio of the ethereal to the material kinetic energy obtaining through all but an infinitesimal fraction of the space will then be $2\mathrm{M}'_1/\mathrm{N}'$, a quantity which cannot be infinite and may be very small. Jeans has illustrated this by calculating the value of the ratio in a special case.[3]

Consider a mass of gas at atmospheric pressure, and at a temperature of 15° C., contained in a cubical enclosure of edge equal to l centimetres. Since $2\mathrm{N}'$ is the number of co-ordinates

[1] *Dynamical Theory of Gases*, chapter ix.

[2] *Roy. Soc. Proc.*, vol. lxxvi., A, 1905, p. 305.

[3] "On the Partition of Energy between Matter and Ether," *Phil. Mag.*, vol. x., 1905, p. 96.

of the gaseous molecules, and each molecule is specified by 6 co-ordinates, there will be N′/3 molecules in the enclosure, and in this case we have approximately

$$\frac{N'}{3}=4\times 10^{19}l^3 .$$

The normal modes of vibration of the ether are then known, each vibration corresponding to values of the components of electric and magnetic force of the type

$$\begin{matrix}\cos\\ \sin\end{matrix}\left(\frac{p'\pi x}{l}\right)\begin{matrix}\cos\\ \sin\end{matrix}\left(\frac{q'\pi y}{l}\right)\begin{matrix}\cos\\ \sin\end{matrix}\left(\frac{r'\pi z}{l}\right),$$

when p', q', r' are integers.[1]

The frequency p of this vibration is determined by the equation

$$(p'^2+q'^2+r'^2)\frac{\pi^2}{l^2}=\frac{p^2}{c^2},$$

where c is the constant of radiation. We may take the small quantity θ as the mean duration of a collision, and for air at 15° C. this will be of the order of 10^{-13} seconds. We take the quantity M'_1 as representing the number of degrees of freedom of the ether for which the period is greater than 10^{-14} seconds. Now the period determined from the preceding equation is

$$\frac{2\pi}{p}=\frac{2l}{c\sqrt{p'^2+q'^2+r'^2}}.$$

Taking $c=3\times 10^{10}$, the upper limit of $\sqrt{p'^2+q'^2+r'^2}$, in order that the vibration may have a period of not less than 10^{-14} seconds, is found to be 20,000 $l/3$. Moreover, the number of sets of positive integral values of p', q', r', for which

$$\sqrt{p'^2+q'^2+r'^2}<\tau,$$

where τ is large, is approximately $\frac{1}{6}\pi\tau^3$, so that in the present case the number of sets of values of p', q', r' is approximately

$$\frac{4\pi\times 10^{12}}{81}l^3.$$

Each system of values of p', q', r' gives four principal co-ordinates, so that for our present purpose

$$M'_1=\frac{16\pi\times 10^{12}}{81}l^3;$$

[1] See Lord Rayleigh's *Sound*, § 267.

so that, substituting for N′ its approximate value, we find

$$\frac{2M'_1}{N'} = \frac{8\pi}{243 \times 10^7} = 10^{-8}, \text{ approximately.}$$

The energy of the ether is therefore almost inappreciable, however large the enclosure may be.

We see then that the number of vibrations whose frequencies lie within any given range of values will vary as the cube of the linear dimensions, that is to say, as the volume, in the case of a cubical enclosure, and a similar proof will hold in the case of a spherical or cylindrical boundary. We may therefore take it generally that the total energy of the ether will be proportional to the volume, so that the energy per unit volume, and therefore the intensity of radiation at a given temperature, measured by the amount of ether energy per unit volume, will be independent of the dimensions of the enclosure. Jeans remarks that "it is also easy to see that it will be independent of the shape of the vessel, although less easy to see how to construct a satisfactory formal proof of this statement."

We now see that we may, in a certain sense, speak of the temperature of the ether within the enclosure. If the enclosed material system consist of a number of separate bodies, we know that the radiation within the enclosure will tend to equalise their temperatures, and it appears natural to regard the process as one analogous to the conduction of heat, and the ether in the final state as possessing the temperature of the bodies. Looking at the question from the point of view of statistical mechanics, and, as a first approximation, treating as invariable the energies of the high frequency vibrations, and, if necessary, also the rapid vibrations of the material systems, we see that the representative point in the ideal space of Appendix K will be constrained to move in a compartment in which the only variable co-ordinates are those corresponding to the low frequency ether vibrations, and certain of the co-ordinates of the material systems, these latter always including those which determine the translatory motions of the matter, that is to say, those which determine the temperatures of the material bodies. Throughout all except an infinitesimal fraction of the compartment, the mean energies of the degrees of freedom of the matter and the ether are equal, so that there is a tendency to the equalisation, not only of the temperatures of the material bodies, but also of the temperatures of the matter and the ether.

The temperature of the ether has not, however, the definite meaning possessed by the temperature of matter. When the latter is known, the kinetic energy per unit mass is determined,

but it is otherwise in the case of the ether; for, although we know the energy of each of the slower modes of vibration, we do not know how many of these modes have to be included in the total energy of the ether. Modes of vibrations of much higher frequency have to be included when the ether is in equilibrium with a mass of hydrogen gas, than when this is replaced by mercury vapour, the molecular movements of which are much slower. We see therefore that the intensity of radiation depends on the nature, as well as on the temperature, of the matter with which it is in equilibrium.

Now let one or more masses of gas be left to themselves in undisturbed ether, and suppose the original total energy to be entirely that of the principal degrees of freedom corresponding to the normal modes of vibration. The first process will be a very rapid transfer of energy between the different degrees of freedom of the gas, resulting, in the course of a small fraction of a second, in a distribution of velocities in accordance with Maxwell's law.

If the masses of gas are small, the next stage will be the equalisation of temperature by conduction through each mass of gas. Simultaneously with this there will be a transfer of energy between the principal degrees of freedom of the molecules and the ether vibrations of low frequency. The temperatures of the gaseous masses will thus become equalised, while the ether will acquire an amount of energy equal to that of a finite number of molecules of the gas, but small in comparison with the total material energy. The time occupied by these changes will be measured in minutes, days, or centuries, according to the linear scale of the system.

Finally, a transfer will begin to become sensible between the energy of the principal degrees of freedom of the gas and that of degrees of freedom which may be either in the ether or the atoms of the gas, but which have the common characteristic that they represent vibrations of high frequency. This third period will, unless the gas be very hot, be measured in millions, or even billions, of years. In a finite system enclosed within a perfectly reflecting boundary impervious to energy the energy thus liberated from the matter will accumulate in the ether.

These considerations show that the conception of the radiation corresponding to a given temperature is not a definite concept,[1] for it depends, not only on the temperatures of the bodies within the enclosure considered, but also on the dimensions of the enclosure and on the time which has elapsed since the system was enclosed. We saw, for example, in Appendix K, that when the

[1] Jeans, *Roy. Soc. Proc.*, vol. lxxvi., A, 1905, p. 306.

system has been enclosed for an infinite time, the radiation corresponding to any temperature is infinite—a result which can only be reconciled with the finiteness of the total energy within the enclosure by the temperature vanishing.

Since we cannot practically form an enclosure impervious to energy, it becomes important to consider what modifications are required when the boundary, instead of being perfectly reflecting, is simply the most perfect attainable in practice. Suppose then, as before, that when the system is first enclosed the whole of the energy resides in the matter. Let E and E_0 be the energy per unit volume of ether vibrations of frequency p within and outside the enclosure, respectively, and assume that leakage takes place across the boundary at the rate $k(E-E_0)$, where k is a constant. Let θ be a small time defined as before, then when p is large, the rate at which ether vibration of high frequency gains energy will involve a factor $e^{-p\theta}$, so that the rate of increase of E will be determined by the equation

$$\frac{dE}{dt} = Ae^{-p\theta} - k(E-E_0),$$

the solution of which is

$$E = (E_0 + Ak^{-1}e^{-p\theta})(1-e^{-kt}),$$

so that ultimately

$$E = E_0 + Ak^{-1}e^{-p\theta}.$$

We find experimentally that the value of E_0 must be very small, while the value of the second term of the right-hand member is necessarily small, since p is large. The energy of high frequency ether vibrations therefore tends towards a limit which is small in comparison with that predicted by the equipartition theorem.

The total energy per unit volume of all the vibrations of high frequency say ΣE, tends to the value

$$\Sigma E_0 + \Sigma Ak^{-1}e^{-p\theta}.$$

Now ΣE_0, the energy per unit volume of these vibrations outside the enclosure, is necessarily finite, and is actually known to be small. The number of modes of high frequency vibration of the ether inside any given enclosure, and which lie between two near values of p, p and $p+dp$, will be proportional to p^2dp; for, according to the calculation on page 581, the number of modes of vibration of the ether, of which the frequency is less than p, inside a cube of edge l, is

$$\frac{2}{3}\frac{l^3p^3}{\pi^2c^2},$$

therefore the number of modes of vibration of frequency between p and $p+dp$ is

$$\frac{2l^3}{\pi^2 c^2} p^2 dp .$$

Since the integral

$$\int^{\infty} e^{-p\theta} p^2 dp$$

is convergent, it follows that the total energy of these high frequency vibrations tends towards a finite limit, so that instead of the energy of the ether tending to become infinite in comparison with that of the matter, the two tend towards a finite ratio. Neither of the actual energies can be permanent, since the system inside the enclosure is not a conservative one; but this definite ratio between the energies of the ether and matter gives meaning to the expression *radiation at a given temperature* as long as we consider the same enclosure and the same enclosed matter.

At the absolute temperature T each degree of freedom of the ether within the enclosure containing gas possesses energy of amount $\frac{1}{2}$RT, so that the energy per cubic centimetre of vibrations of frequencies between p and $p+dp$ is

$$\frac{\mathrm{RT}}{\pi^2 c^3} p^2 dp ,$$

or, in terms of the wave-length λ in free ether, the energy of wave-lengths between λ and $\lambda + d\lambda$ is

$$8\pi \mathrm{RT} \lambda^{-4} d\lambda ,$$

and therefore the total energy of radiation at the temperature T is

$$8\pi \mathrm{RT} \int_{\lambda_0}^{\infty} \lambda^{-4} d\lambda + \int_0^{\lambda_0} f(\lambda, \mathrm{T}, t) d\lambda ,$$

in which λ_0 is the shortest wave-length for which the vibrations may be supposed to possess their full energy, and the second integral represents the energy of waves of lengths less than λ_0, the energy of radiation of which is a function of t, the time which has elapsed since the closing in of the ether, as well as of λ and T.

Now suppose several bodies, each at its own temperature, to be introduced into a perfectly reflecting enclosure. It is then usually assumed that, after the lapse of an infinite time, equalisation of the temperature would take place, as a consequence of the second law of thermodynamics. Now, as we have seen, Jeans has shown this to be inapplicable, but the argument on page 295 in support

of this position, depending on an assumption which still remains to be demonstrated, suggests the possibility of equalisation taking place, in consequence of the fact that the charge of an electron remains invariable, whatever the nature of the matter of which it is a constituent. It will presently be shown that this is the case, and in the meantime we shall assume that equalisation does take place, and that when the final state is attained, the amount of energy in the ether, and its distribution amongst different wave-lengths, remains constant. The energy per unit volume of radiation of wave-lengths lying between λ and $\lambda + d\lambda$ must then be expressible in the form $F(\lambda, T)d\lambda$.

The positions of the bodies within the enclosure being immaterial, we will assume, with Jeans,[1] that they are spread over the walls of the enclosure so as to completely cover every part of it. The whole system within the reflecting walls will still remain impervious to energy. Now suppose a minute hole to be pierced in the walls, and the flow of radiation energy through it to be maintained constant by supplying energy of appropriate amount and constitution to the interior of the enclosure. The issuing stream of energy will then remain constant, and its constitution will therefore be represented by $F(\lambda, T)d\lambda$.

Let these ideal conditions now be modified into experimentally possible ones by substituting for the ideal perfectly reflecting walls, an actual material enclosure, the walls of which are maintained at a uniform temperature T. The stream of energy will again be steady, and is found experimentally to be independent of the nature and reflecting power of the walls. The constitution of the issuing stream of energy will therefore be expressible in the form $f(T, \lambda)d\lambda$. It is usually assumed that the constitution of the energy would be the same in each case, which would make F and f identical; and this is the assumption tacitly involved in the common employment of the expression "radiation at a given temperature," the validity of which has already been discussed. Now we have already shown, by means of the equipartition theorem, that for the gas within the enclosure

$$F(\lambda, T) = 8\pi RT\lambda^{-4}.$$

If therefore it could be shown that f is identical with F, it would prove the applicability of the partition theorem to the energy within the substance of the walls as well as to the energy within the enclosure. Jeans, in the paper last referred to, points out that, for long waves, experiment appears to suggest the identity of f and

[1] *Phil. Mag.*, vol. xvii., 1909, p. 230.

F,[1] and in his paper on "The Motions of Electrons in Solids," he shows, as we shall presently see, that this identity, for vibrations of great wave-length, is a necessary consequence of the electron theory.

It was pointed out on page 293[2] that if charges e, e', etc., move with velocities $\mathbf{u}$, $\mathbf{u}'$, etc., there will be no radiation of wave-lengths which are large compared with the distances between the charges if $\Sigma e\dot{\mathbf{u}} = 0$. If this condition be not satisfied, let $\mathrm{E}\mathbf{U} = \Sigma e\mathbf{u}$; then

$$-\mathrm{E}\dot{\mathbf{U}} + \Sigma e\dot{\mathbf{u}} = 0,$$

so that there is no resultant radiation from E, e, e', etc. The radiation from the charge E moving with the velocity $\mathbf{U}$ is therefore identical with the radiation from the charges e, e', etc. moving with their own proper velocities. Now consider an element of volume of small linear dimensions compared with the wave-lengths we are concerned with, and containing a large number $\mathrm{N}dv$ of electrons, the velocity components of which are u, v, w; u', v', w', etc., their average values being u_0, v_0, w_0. Then the radiation from these electrons can be replaced by that from a single electron of charge $\mathrm{N}edv$, and having the velocity components u_0, v_0, w_0. If f is the acceleration of this electron and μ is the magnetic permeability of the medium, that is to say, of the conductor in question, then, where $\mathrm{V}^2 = c^2/\mathrm{K}\mu$, the radiation from this electron in time t is[3]

$$\int_0^t \frac{2\mu}{3\mathrm{V}}(\mathrm{N}edvf)dt.$$

The components of the total current C are

$$\mathrm{N}edvu_0, \quad \mathrm{N}edvv_0, \quad \mathrm{N}edvw_0,$$

so that

$$\mathrm{N}edvf = \dot{\mathrm{C}},$$

and the radiation in time t is

$$\frac{2\mu}{3\mathrm{V}}\int_0^t \dot{\mathrm{C}}^2 dt.$$

[1] See Planck, *Vorlesungen über Wärmestrahlungen*, 1906, p. 157; and Lorentz, *Amsterdam Proc.*, 1903, p. 678.

[2] See also *Æther and Matter*, p. 225.

[3] See *Æther and Matter*, p. 227. Since the radiation is obtained by integrating over a sphere of large radius from the interior of which all absorbing and dispersing electrons are removed, V is the velocity in the medium when the latter is freed from absorption and dispersion, and is given by the equation $\mathrm{V}^2 = c^2/\mathrm{K}\mu$. For $p=0$ it becomes identical with the actual velocity in the medium.

Now we may expand $\dot{C}$ in the form

$$\dot{C} = \frac{1}{\pi}\int_0^\infty (A_p \cos pt + B_p \sin pt)dp$$

where

$$A_p = \int_0^t \dot{C} \cos pt dt, \quad B_p = \int_0^t \dot{C} \sin pt dt.$$

Then [1]

$$\int_0^t \dot{C}^2 dt = \frac{1}{\pi}\int_0^\infty (A_p^2 + B_p^2)dp,$$

so that the total emission of the element dv in time t is

$$\frac{2\mu}{3\pi V}\int_0^\infty (A_p^2 + B_p^2)dp \qquad . \quad . \quad . \quad (12)$$

and the coefficient of dp in this expression represents the emission of radiation of frequencies lying between p and $p + dp$.

Again,

$$A_p = \int_0^t \dot{C} \cos pt dt = \Big[C \cos pt \Big]_0^t + p\int_0^t C \sin pt dt;$$

and when t is large, the first term may be neglected, so that we may write

$$A_p = p\int_0^t C \sin pt dt, \quad B_p = -p\int_0^t C \cos pt dt.$$

Then, where C_1, C_2 are the values of C at any instants t_1, t_2 within the interval from 0 to t,

$$A_p = p\int_0^t C_1 \sin pt_1 dt_1 = p\int_0^t C_2 \sin pt_2 dt_2,$$

and therefore

$$A_p^2 = p^2 \int_0^t\int C_1 C_2 \sin pt_1 \sin pt_2 dt_1 dt_2.$$

[1] Lord Rayleigh, *Phil. Mag.*, vol. xxvii. (1889), p. 406, or *Scientific Papers*, vol. iii. p. 268.

Similarly,

$$B_p^2 = p^2 \int_0^t \int_0^t C_1 C_2 \cos pt_1 \cos pt_2 dt_1 dt_2 .$$

Therefore

$$A_p^2 + B_p^2 = p^2 \int_0^t \int_0^t C_1 C_2 \cos p(t_1 - t_2) dt_1 dt_2$$

$$= p^2 \int_{t_1=0}^{t_1=t} \int_{t_2=0}^{t_2=t} C_1 C_2 \cos p(t_1 - t_2) d(t_1 - t_2) d\left(\frac{t_1 + t_2}{2}\right).$$

Let the two instants t_1, t_2 be at equal intervals θ from their middle instant t', of which the value is $(t_1 + t_2)/2$, and let C′ be the current at t'. The expectation of change of C is by (5),

$$\dot{C} = -\frac{Ne^2}{m\kappa} C,$$

the integral of which is

$$C = C_0 e^{-\epsilon t},$$

where $\epsilon = Ne^2/m\kappa$. Therefore the current C_2, at an interval θ after it is C′, is given by the equation

$$C_2 = C' e^{-\epsilon\theta} + j,$$

where j is a quantity of which the expectation is zero.

Similarly, since the motion is reversible, the value of C_1 is

$$C_1 = C' e^{-\epsilon\theta} + j',$$

where j' has zero expectation and has no correlation with j.

Substituting these values, we have

$$A_p^2 + B_p^2 = p^2 \int\int C'^2 e^{-2\epsilon\theta} \cos 2p\theta d(2\theta) dt',$$

the terms in j, j' and jj' being omitted since, their average value being zero, they vanish on integration. The exponential $e^{-2\epsilon\theta}$ diminishes very rapidly as the interval 2θ between t_1 and t_2 increases, this being the mathematical expression of the fact that there is very little correlation between the values of C at intervals of time far apart. We may therefore integrate with respect to θ from $-\infty$ to $+\infty$, taking θ always positive in the exponential $e^{-2\epsilon\theta}$, and therefore have

$$A_p^2 + B_p^2 = p^2 \int_0^t C'^2 \frac{2\epsilon}{\epsilon^2 + p^2} dt'.$$

Now

$$C_x = e(u + u' + \; . \; . \; . \; .),$$

so that

$$C_x^2 = e^2(u^2 + u'^2 + \; . \; . \; . \; .) = N dv e^2 \frac{RT}{m},$$

and therefore

$$C^2 = 3RT\frac{Ne^2}{m}dv.$$

Hence

$$A_p^2 + B_p^2 = \frac{2p^2\epsilon}{\epsilon^2 + p^2} 3RT\frac{Ne^2}{m}tdv.$$

Substituting for ϵ its value, we find from (12) that the total emission of the element dv in time t is

$$\left(\int \frac{4\mu}{\pi V} \frac{\kappa}{1 + \frac{\kappa^2 p^2 m^2}{N^2 e^4}} p^2 RT dp\right) tdv \qquad . \qquad . \quad (13)$$

The partition of radiant energy in the matter in the steady state can now be determined by equating the emission to the corresponding absorption. To obtain an expression for the latter, let E be the energy per unit volume of the stream of radiant energy, and C_x the current parallel to the electric force, of numerical value X, in the stream; then the energy absorbed per unit volume per unit time is

$$XC_x = X_2 k',$$

where k' is the conductivity of the substance appropriate to the character of X.

Now, if K is the dielectric constant of the substance,

$$E = \frac{KX}{4\pi c^2},$$

and therefore

$$X^2 = \frac{4\pi c^2}{K}E,$$

and the rate of absorption of energy per unit of volume is equal to

$$k'\frac{4\pi c^2}{K}E.$$

For currents of frequency p we must replace k' by

$$k = \frac{\kappa}{1 + \frac{\kappa^2 p^2 m^2}{N^2 e^4}},$$

and we may write

$$E=\int E_p dp.$$

The absorption of energy in the element dv in time t is therefore

$$\left(\int \frac{4\pi c^2}{K}\frac{\kappa}{1+\frac{\kappa^2 p^2 m^2}{N^2 e^4}}E_p dp\right)tdv.$$

Equating this expression to (13), we have

$$E_p=\frac{p^2 RTK\mu}{\pi^2 c^2 V}=\frac{p^2 RT}{\pi^2 V^3}.$$

E_p therefore depends on the structure of the medium only through the factor $1/V_1{}^3$ as is required by (11), in order to accord with the observed fact that the radiation in a cavity is independent of the nature of the matter enclosing it. It follows that for the radiation in the cavity

$$E_p=\frac{p^2 RT}{\pi^2 V_1{}^3},$$

where V_1 is the velocity in the cavity.

Now if λ is the wave-length of radiation of frequency p, the values of $p\lambda$ will be $2\pi V$ in the metal and $2\pi V_1$ in the cavity. In each case, therefore, the radiation is given by the equation

$$\int E_p dp=\int 8\pi RT\lambda^{-4}d\lambda,$$

which (p. 585) expresses the condition that the partition of energy should be that required by the law of equipartition. The distribution of the energy is therefore in accordance with this law both in the metal and the cavity. That is to say, as is stated on page 325, the energy in a solid conductor is equally divided amongst the vibrations of great wave-length, both in a cavity within it, and throughout its substance, so that each vibration has the energy appropriate to an absolute temperature T determined by the kinetic energy of the electrons. This must therefore hold good also for the emitted radiation, as it follows that this is simply that part of the energy of the electrons which is yielded up to the ether.

It can now be shown that the laws of Stefan and Wien are approximately fulfilled.[1]

[1] J. H. Jeans, "On the Laws of Radiation," *Roy. Soc. Proc.*, vol. lxxvi., A, 1905, p. 547.

If the boundary of the radiator is a semi-infinite plane, and θ is a coefficient of extinction, the energy generated at a point at a distance r from some specified small area on the boundary will be reduced, in crossing the boundary through this small area, in the ratio $e^{-\theta r}$. If, therefore, energy is transferred to the ether at the uniform rate G per unit volume per unit time throughout the radiator, the stream of energy crossing unit area of the boundary will be

$$\int_0^\infty \frac{G}{4\pi r^2} e^{-\theta r^2} \pi r^2 dr = \frac{G}{4\theta},$$

which is a definite finite quantity. Since the contribution corresponding to large values of r is infinitesimal, the stream crossing the boundary at any point may be considered as proceeding only from the portions of the radiator in the immediate vicinity of the point. The stream will therefore depend on the structure of the surface-layers only, and not on the shape and size of the radiator.

Suppose that at any point of the surface, at temperature T, the energy of the issuing radiation of wave-lengths lying between λ and $\lambda + d\lambda$ is

$$\phi(\lambda, T)d\lambda$$

per unit volume. The function ϕ, in addition to depending on λ and T, will also depend on :— c, the constant of radiation; e, the charge of an electron; m, the mass of an electron; R, the constant of the kinetic theory of gases, this being such that the mean kinetic energy of an electron is $\frac{3}{2}RT$; K, the specific inductive capacity of the ether, measured in whatever units are in use; k_1, k_2, k_3, ... , quantities specifying the structure of the radiator, *e.g.* the number of free electrons per unit volume, the shape, size, mass, etc., of the atoms and molecules. The law of radiation can then be expressed more completely in the form

$$\phi(\lambda, T, c, e, m, R, K, k_1, k_2, k_3 \ . \ . \ . \ .)d\lambda .$$

The values of k_1, k_2, k_3, ... , for which the function $\int\phi d\lambda$ is a maximum, can then be obtained by solving the equations

$$\frac{\partial}{\partial k_1}\int \phi d\lambda = 0, \quad \frac{\partial}{\partial k_2}\int \phi d\lambda = 0, \text{ etc.,}$$

thus determining the properties of the solid such that the total radiation at a given temperature is a maximum, and which may

be called the *radiator of maximum efficiency* for the given temperature. Let ϕ_m be the value of ϕ for this radiator. Then, by eliminating $k_1, k_2, k_3, \ldots$ by means of the foregoing equations, ϕ_m can be expressed as a function of the remaining quantities. In terms of the units, L of length, M of mass, t of time, K of specific inductive capacity, and T of degrees of temperature, the physical dimensions of these quantities are as follows:—

Quantity . .	λ	T	c	e	m	R	K
Dimensions .	L	T	Lt^{-1}	$L^{\frac{3}{2}}M^{\frac{1}{2}}t^{-1}K^{\frac{1}{2}}$	M	$LMt^{-2}T^{-1}$	K

Here there are seven quantities expressed in terms of five independent physical units. The quantities may therefore be combined in two independent ways so as to form a mere number, and these we may take to be

$$a_1 \equiv RTm^{-1}c^{-2}, \quad a_2 \equiv \lambda RTKe^{-2}.$$

Any other pure number which it is possible to form from these seven quantities must then be of the form $f(a_1, a_2)$.

Now the dimensions of ϕ_m are $L^{-3}Mt^{-2}$, viz. those of energy per unit volume per unit wave-length, and these are the dimensions of $\lambda^{-4}RT$. The ratio of ϕ_m to this quantity is, therefore, a pure number, so that ϕ_m must be expressible in the form

$$\phi_m = \lambda^{-4}RTf(a_1, a_2).$$

The mean kinetic energy of a free electron at temperature T is $\frac{3}{2}RT$, so that the value of $\bar{v}^2$, the mean square of its velocity, is $\frac{3}{2}RTm^{-1}$, and therefore,

$$\frac{3}{2}a_1 = \frac{\bar{v}^2}{c^2}.$$

But at 100° C. the value of $\bar{v}$ is 7×10^6 centimetres per second,[1] while $c = 3 \times 10^{10}$. The value of a_1 is therefore $3{\cdot}6 \times 10^{-8}$, a quantity small enough to be neglected. Assuming that when a_1 is put equal to zero, the function $f(a_1, a_2)$ tends to a definite limit $f(0, a_2)$; it will follow, since the actual value of a_1 is very small, that ϕ_m may be very approximately expressed in the form

$$\phi_m = \lambda^{-4}RTf(a_2),$$

or, substituting for a_2 its value, and dropping the universal constants R, T, K, and e,

$$\phi_m = \lambda^{-4}Tf(\lambda T),$$

[1] See Jeans' *Dynamical Theory of Gases*, p. 209.

which will therefore give the law of radiation from a radiator of maximum efficiency for temperature T and wave-length λ. This is a purely ideal body which may be designated a *perfect radiator*. Integrating the above expression with respect to λ, we find that the total radiation is of the form σT^4, where

$$\sigma = \int_0^\infty x^{-4} f(x) dx,$$

which is the expression of Stefan's law.

Let $\lambda_{\text{max.}}$ denote the wave-length for which the energy per unit wave-length, that is to say, the coefficient of $d\lambda$ in the expression for the law of radiation, is a maximum, then $\lambda_{\text{max.}}T$ must be a root of the equation

$$\frac{\partial}{\partial x}\{x^{-4}f(x)\} = 0,$$

so that

$$\lambda_{\text{max.}} = a,$$

where a is a constant, which is the expression of Wien's law.

The constant σ is a function only of c, e, m, R and K, and is of dimensions $L^{-2}Mt^{-2}T^{-4}$, and the only way in which these quantities can be combined to form a quantity of the dimensions of σ is through an expression of the form $e^{-6}R^4K^3$. Jeans draws the inference that σ must be of the same order of magnitude as the latter expression, and similarly that the constant a of Wien's law must be of the same order of magnitude as $e^2R^{-1}K^{-1}$. By comparison of these expressions with the observed values

$$\sigma = 5{\cdot}2 \times 10^{-5} \text{ (Kurlbaum)}$$
$$a = 0{\cdot}295 \text{ (Lummer and Pringsheim)},$$

and taking $R = 9{\cdot}3 \times 10^{-17}$, the first approximate equality

$$e^{-6}R^4K^3 = 5{\cdot}32 \times 10^{-5}$$

gives

$$eK^{-\frac{1}{2}} = 1{\cdot}8 \times 10^{-10},$$

while from the second relation

$$e^2R^{-1}K^{-1} = 0{\cdot}294$$

it follows that

$$eK^{-\frac{1}{2}} = 51 \times 10^{-10}.$$

The experimental value of $eK^{-\frac{1}{2}}$, as determined by Sir J. J. Thomson, is 3×10^{-10}.

Jeans remarks that the difference between this and the values obtained above is not greater than is to be expected in view of the roughness of the method, for, as he points out, the calculations would have been the same if e had been expressed in terms of Heaviside's rational units (see Appendix H), in which case Thomson's experimental value of $eK^{-\frac{1}{2}}$ would have been 4π times as great, that is to say, about 38×10^{-10}.

APPENDIX N.

ASTRONOMICAL ANOMALIES AND THE GENERALISATION OF NEWTON'S LAW OF GRAVITATION.

The electrical theory of matter suggests the possibility, as was pointed out on page 421, that the gravitational attraction between two bodies may vary to some extent with the magnitudes and directions of their velocities, as well as with their masses and positions. A brief consideration of some of the attempts which have been made to account for the astronomical anomalies referred to in Chapter XXIII., on the basis of Weber's earlier electrical theory, will therefore be of interest.

Weber, in his development, referred to on page 39, of Ampère's theory of electric currents, proposed a hypothetical law for the numerical value E of the electrical force E between two moving charges, which would lead to the results observed by Ampère. This law, expressed in terms of Weber's electrodynamic system of units, is represented by the equation

$$\mathrm{E}=\frac{ee'}{r^2}(1-a^2\dot{r}^2+2a^2r\ddot{r}),$$

where e, e' are the amounts of the charges, r is the distance between them, and a is the reciprocal of a velocity, its value being $1/c\sqrt{2}$. The potential energy of the two particles would therefore be given by the expression

$$\frac{ee'}{r}(1-a^2\dot{r}^2).$$

Assuming matter to be built up of electric charges, this would give for the gravitation potential of two masses m, m' at a distance r from each other

$$\mathrm{V}=\mathrm{G}\frac{mm'}{r^2}\left(1-\frac{\dot{r}^2}{c^2}\right)$$

where G is a constant.

This formula has been applied to the general investigation of planetary motions by C. Seegers[1] and G. Holtzmüller,[2] and numerically by F. Tisserand[3] and H. Servus,[4] with the result that the anomaly in the motion of the perihelion of mercury, as determined by this formula, was found to amount in a century to fourteen seconds, its observed value, as stated on page 418, being about forty-one seconds. In order to account for the whole of the anomaly it would be necessary to modify Weber's formula by replacing c by a suitable smaller constant. This, however, would make the velocity of propagation of gravitational action considerably less than that of light.

P. Gerber[5] approaches the problem by regarding the gravitational potential as something propagated from the attracting mass m_1 to the attracted mass m_2 with a proper velocity of its own, v, to which is to be added the velocity of m_1 relative to m_2. Suppose these to be at a distance r at the time t, and to be approaching each other with a velocity which is small compared with v. When the bodies are not in relative motion the potential will be

$$V_0 = \frac{m_1 m_2}{r}.$$

When m_1 and m_2 are approaching each other, then, if Δt be the time taken by the potential emitted from m_1 to reach m_2, we shall have $r = v\Delta t$, and therefore the distance traversed will be

$$r - \Delta r = \Delta t\left(v - \frac{\Delta r}{\Delta t}\right) = r\left(1 - \frac{1}{v}\frac{\Delta r}{\Delta t}\right).$$

The amount of transmitted potential will be inversely proportional to this distance, and the relative velocity of transmission is $v - \Delta r/\Delta t$, so that the potential V will be proportional to

$$\frac{v}{v - \dfrac{\Delta r}{\Delta t}},$$

and therefore

$$V = \frac{m_1 m_2}{r\left(1 - \dfrac{1}{v}\dfrac{\Delta r}{\Delta t}\right)^2};$$

1 *Goettingen Dissertation*, 1864.
2 *Zeitschr. Math. Phys.*, 1870, p. 69.
3 *Comptes Rendus*, 1872, p. 760; 1890, p. 313.
4 *Halle Dissertation*, 1885.
5 *Zeitschr. Math. Phys.*, vol. xliii., 1898, p. 93.

or, if the velocity of approach be sufficiently small compared with v,

$$V = \frac{m_1 m_2}{r\left(1 - \frac{\dot{r}}{v}\right)^2} = \frac{m_1 m_2}{r}\left(1 + \frac{2\dot{r}}{v} + \frac{3\dot{r}^2}{v^2}\right).$$

Gerber finds that the anomaly of forty-one seconds will be completely accounted for by taking $v = 305500$ kilometres per second, that is to say, within the limits of errors of observation, by taking $v = c$, the velocity of radiation in the ether. For the perihelion motion of Venus this value introduces an anomaly of about eight seconds per century, but does not lead to any difficulties in the cases of the other planets, the amounts of the anomaly in the motion of the perihelion being:—for the Earth, 3·6 seconds per century; for the Moon, ·06 sec.; for Mars, 1·3 secs.; for Jupiter, ·06 sec.; for Saturn, ·01 sec.; for Uranus, ·002 sec.; and for Neptune, ·0007 sec. per century.

This investigation is not one that suggests any physical representation of gravitational action, and is only presented as a preliminary inquiry into some of the conditions to be satisfied by a satisfactory theory, and as indicating the possibility of gravitational propagation with the velocity of radiation. It could not, however, with such a velocity, be of the nature of radiation, as the impossibility of this was shown in Chapter XXIII.

APPENDIX O.[1]

THE CLUSTERING OF GRAVITATIONAL MATTER IN THE UNIVERSE.

THE mean density of ponderable matter through any very large spherical volume of space, supposed to contain only positive masses, is smaller, the greater the radius; and is infinitely small for an infinitely great radius. For if this were not the case, a majority of the bodies in the universe would each experience infinitely great gravitational force.

Let V be the volume of a portion of space contained within a closed surface S, without and within which there are ponderable bodies. Let M be the total amount of ponderable matter contained within S, and ρ its mean density throughout the volume V, so that

$$M = \rho V.$$

Let Q be the mean value of the normal component of the gravitational force at all points of S. Then, by Green's theorem,

$$QS = 4\pi M = 4\pi\rho V.$$

If the surface S be a sphere of radius r,

$$S = 4\pi r^2, \; V = \frac{4\pi}{3} r^3 = \frac{1}{3} rS,$$

so that

$$Q = \frac{4\pi}{3} r\rho = \frac{M}{r^2}.$$

If, therefore, ρ is finite and r is infinitely great, the mean value of the normal component force over the spherical surface is infinitely great.

[1] Lord Kelvin's *Baltimore Lectures*, pp. 270–271.

If E is the gravitational energy, and R the gravitational force at the point x, y, x, then[1]

$$E = \frac{1}{8\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} R^2 dxdydz,$$

and when a vast number, N, of equal masses come from rest at infinite distances from one another to an equably spaced distribution throughout a sphere of radius r, the exhaustion of the gravitational energy is found to be $3/10 \cdot Fr$, where F is the resultant attraction of all of them on a material point, of mass equal to the sum of their masses, situated on the surface of the sphere. Now this exhaustion is entirely expended in generating kinetic energy, so that

$$\sum \frac{1}{2} mv^2 = \frac{3}{10} Fr.$$

But the sun's mass is 324,000 times that of the earth, and therefore the whole quantity of matter within the sphere is $3{\cdot}24 \times 10^{14}$ times the earth's mass. Therefore, if g be the value of terrestrial gravity,

$$Q = 3{\cdot}24 \times 10^{14}\left(\frac{6{\cdot}37 \times 10^8}{3{\cdot}09 \times 10^{16}}\right)^2 g = 1{\cdot}37 \times 10^{-11} g;$$

or, taking g as $\cdot 01$ kilometre per second per second,

$$Q = 1{\cdot}37 \times 10^{-13} \text{ kilometre per second per second,}$$

so that

$$F = 1{\cdot}37 \times 10^{-13} \Sigma m,$$

and therefore

$$\frac{\Sigma mv^2}{\Sigma m} = \frac{3}{5} \times 1{\cdot}37 \times 10^{-13};$$

or, substituting for r its value, and taking the masses equal,

$$\sqrt{\frac{\Sigma v^2}{N}} = \sqrt{\left(\frac{3}{5} \times 1{\cdot}37 \times 10^{-13} \times 3{\cdot}09 \times 10^{16}\right)} = 50{\cdot}4 \text{ kilometres per}$$
second.

[1] Thomson and Tait's *Natural Philosophy*, § 549.

APPENDIX P.

HYDRODYNAMICAL ETHER THEORIES OF GRAVITATION AND THE DETERMINATION OF THE SCALE OF ATOMIC STRUCTURE.

THE similarity of the lines of force in a gravitational field containing two particles to those of a magnetic field containing two magnetic poles, which was pointed out by Maxwell[1] (p. 423), is obviously a necessary consequence of the similarity of the observed law of force between attracting particles and magnetic poles respectively. It was stated in Chapter XXIII. that fields of force capable of representing the observed phenomena of gravitation could be produced either by a continuous flow of the fluid into and out of the material molecules, or by an alternating flow accompanied by pulsations of the molecules. The first alternative, being the simpler one, will be considered first.

Let us consider, as suggested by Lord Kelvin (p. 440), a number of long open tubes, so narrow as not sensibly to impede the external motion of the fluid in which they are immersed, and suppose a twistless flow to be established through them. Now the twistless flow of a fluid within a simply connected region is characterised by the existence of a single-valued velocity-potential. Also, as was first shown by Lord Kelvin,[2] in a perfect fluid whose density is uniform or a function of the pressure only, and is subject to forces having a single-valued potential, the circulation in any circuit moving with the fluid is constant. The extremities of the tubes will therefore, as regards external space, act as constant sources or sinks, that is to say, points from or towards which there is a constant outflow or inflow which is equal in all radial directions. Let the strength of a source, that is to say, the total outward flux across a small closed surface surrounding it, be denoted by σ. Then, if the fluid be supposed

[1] "A Dynamical Theory of the Electromagnetic Field, Gravitational Note," *Phil. Trans.*, vol. clv., 1865; *Scientific Papers*, vol. i. p. 570.

[2] "On Vortex Motion," *Edin. Trans.*, vol. xxv., 1869, p. 217; or Lamb's *Hydrodynamics*, art. 33.

to be at rest at infinity, the value of the velocity-potential ϕ at any point will be a function only of r, the distance of the point from the source. Therefore, for a single source,

$$\phi = -\frac{\sigma}{4\pi r},$$

for this gives a radial flow from the source; and if $dS = r^2 d\omega$ be an element of a spherical surface having its centre at the source, we have, since $\partial\phi/\partial r = \sigma/4\pi r^2$,

$$-\iint \frac{\partial \phi}{\partial r} dS = \sigma .$$

Let ρ be the density of the fluid, then the energy arising from any distribution of sources and sinks σ_1, σ_2, etc., will be given, so far as it depends on their relative configuration, by the integral

$$\tfrac{1}{2}\rho \iint \phi \frac{\partial \phi}{\partial n} dS$$

taken over a system of small closed surfaces surrounding the several sources and sinks, the normal n being measured outwards for a source, and inwards for a sink.

Let the surface S, fig. 35, represent the outer boundary of the fluid, and let two extremely small closed surfaces A_1 and A_2 be supposed to enclose the sources σ_1 and σ_2. Let P be any point in the fluid, and let the velocity potentials at P, due to the general body of fluid, and the sources A_1 and A_2, be ϕ, ϕ_1, ϕ_2, respectively. Then if T be the kinetic energy of the whole field included within the external boundary S and the internal boundaries formed by the small closed surfaces A_1 and A_2, we shall have

$$2T = \rho \int \left(\phi + \phi_1 + \phi_2\right)\left(\frac{\partial \phi}{\partial n} + \frac{\partial \phi_1}{\partial n} + \frac{\partial \phi_2}{\partial n}\right) dS,$$

taken over the surfaces S, A_1, and A_2, dn being measured outwards over S and inwards over A and B. If S be a sphere of immense radius R, with A_1 and A_2 near its centre, the integral over S will be

$$\int \left(\phi - \frac{\sigma_1}{4\pi R} - \frac{\sigma_2}{4\pi R}\right)\left(\frac{\partial \phi}{\partial n} + \frac{\sigma_1}{4\pi R^2} + \frac{\sigma_2}{4\pi R^2}\right) dS$$

$$= \int \phi \frac{\partial \phi}{\partial n} dS + \frac{\sigma_1 \sigma_2}{4\pi} \int \phi\, d\omega - \frac{\sigma_1 + \sigma_2}{4\pi R} \int \frac{\partial \phi}{\partial n} dS - \frac{(\sigma_1 + \sigma_2)^2}{16\pi^2 R^3} \cdot 4\pi r^2,$$

where $d\omega$ is the conical angle subtended at A or B by the element dS of the surface S.

Of the four terms in the preceding expression, the first is independent of any change in the relative position of A_1 and A_2, the third vanishes, since $\int \frac{\partial \phi}{\partial n} dS = 0$ by the equation of continuity, and the last term vanishes because R is very great. The only portion of the integral arising from the simultaneous presence of A and B is therefore

$$\frac{\sigma_1 + \sigma_2}{4\pi} \int \phi d\omega,$$

which vanishes for $\phi = 0$ at infinity, and even in the absence of this condition, this contribution, due to the simultaneous presence

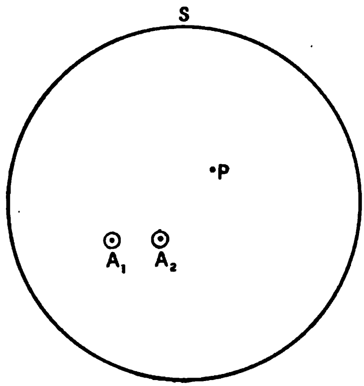


FIG. 35.

of A_1 and A_2, is independent of their positions, and will not be affected by any variation in their relative distance.

The portions of T derived from the integrations over the small surfaces A_1 and A_2 will, similarly, contribute terms depending on the magnitudes of σ_1 and σ_2, and on the values of ϕ at A_1 and at A_2, and the only terms involving the distance r_{12} between A_1 and A_2 will be a term $\sigma_1\sigma_2/4\pi r_{12}$ in each integral. These will contribute $\sigma_1\sigma_2\rho/2\pi r_{12}$ to 2T, or $\sigma_1\sigma_2\rho/4\pi r_{12}$ to T.

If F be the force acting between A_1 and A_2, the amount of work arising from this force, corresponding to a change dr_{12} in their relative distance, will be $F dr_{12}$, and the corresponding change in T will be

$$-\rho \frac{\sigma_1 \sigma_2}{4\pi r_{12}^2} dr_{12}.$$

If, therefore, we assume the work $F dr_{12}$ to be wholly absorbed in giving rise to this change, these two quantities will be numerically equal, so that if F be taken as an attractive force, tending to decrease the distance between A_1 and A_2, we shall have

$$F = \frac{\rho}{4\pi} \cdot \frac{\sigma_1 \sigma_2}{r_{12}^2}.$$

Hence two sources or two sinks attract, and a source and sink repel, each other with a force proportional to the product of their strengths, and varying inversely as the square of the distance separating them, which is in accordance with Newton's law of gravitation, if we suppose the electrons to be such sources or sinks in the ether. This, as was pointed out on p. 441, would necessitate either a slow, steady expansion or contraction of the electrons, or a flow of ether through every electron into or out of a four-dimensional space containing our three-dimensional universe.

We have now to consider the legitimacy of the assumption that the work arising from the force F is wholly absorbed in producing the change in the relative distance of A_1 and A_2. Since the integrals taken over the small surfaces A_1 and A_2 contain the terms $\sigma_1\phi$ and $\sigma_2\phi$, it follows that any change in r_{12} will necessitate changes in the positions of A_1 and A_2, which will give rise to changes in the value of ϕ at A_1 and A_2. Now, in the case of the ether in the universe, the velocity potential is to be attributed entirely to the distribution and the motions of the sources or sinks, or sources and sinks, corresponding to the electrons. If the distances between any two electrons are great in comparison with the magnitudes of the corresponding sources or sinks, the variation in ϕ arising from small changes in their positions will therefore be negligible for the purpose of a first approximation. We know, moreover, that this condition is fulfilled, for the distance between two electrons is known to be great relatively to their dimensions, and by the dimensions of an electron we simply mean its sphere of sensible action.

The source and sink theory therefore leads to Newton's law of gravitation under the circumstances where it is found to be in accordance with observation, but only as an approximation, leaving outstanding small forces depending on the motions of the electrons of the character which appear to be requisite in order to account for the astronomical anomalies referred to in Chapter XXIII. We see also that these outstanding forces, as far at least as their principal terms are concerned, are of opposite sign in a source and a sink. That is to say, negative gravitational mass is provided for.

Let m be the mass of an electron, assumed to act as a source, or a sink, of strength σ; then, where G is the constant of gravitation and r is the distance between the mass-centres of two masses M_1 and M_2 containing n_1 and n_2 electrons respectively, their gravitational attraction will be

$$G\frac{M_1M_2}{r^2}=\frac{\rho n_1n_2\sigma^2}{4\pi r^2}=\frac{\rho M_1M_2\sigma^2}{4\pi m^2r^2},$$

so that

$$G=\frac{\rho\sigma^2}{4\pi m^2},$$

and is therefore no longer a mere number, as was assumed on page 30. Since the mass of an electron is of the order 10^{-27}, and the numerical value of G is $6{\cdot}7\times10^{-8}$, if we assume the density of the ether (p. 185) to be of the order 10^{12}, we find that σ is of the order 10^{-36}.

Dr G. A. Schott[1] has, as mentioned on p. 442, worked out mathematically a theory of expanding electrons, primarily in order to account for the definiteness of atomic structure. Dr Schott states that he finds it would be necessary, in order to account for gravitation on this hypothesis, to assume an ether density of an order exceeding 10^{100}.

He considers the case of a ring of n equidistant electrons moving with uniform velocity v in a circle of radius r, the mass of each electron being m, its charge e, and its mean radius a. Then if T, P, are the tangential and normal components, respectively, of the mechanical force on an electron, and writing $v/c=\beta$, c being the constant of radiation, the equations of motion will be

$$T=\frac{d(mv)}{dt}+\frac{2e^2\beta^2}{3r^2(1-\beta^2)^2},\quad P=\frac{mv^2}{r}.$$

The first term in the first equation would vanish in steady motion if m were constant, but it is necessary to retain it to allow of m varying while v remains constant. The second term represents the reaction of the ether on the electron due to radiation. It is pointed out that the mass of the electron, if this be assumed to be entirely electromagnetic, or in any case the electromagnetic portion of the mass, can be shown by dimensional considerations to be of the form $e^2/c^2a\,.\,\psi(\beta)$, where $\psi(\beta)$ is a function depending only upon β and the shape of the electron. If a be supposed to vary, the variation must be extremely small, since

$$\frac{\dot{m}}{m}=-\frac{\dot{a}}{a},$$

[1] *Phil. Mag.*, vol. xii., 1906, p. 21.

in order to correspond with the observed constancy of mass of the electron and the observed fineness and fixity of the spectrum lines. In the paper 10^{-10} is suggested as a maximum value of $\dot{a}/a$, but in a letter to the author Dr Schott substitutes 10^{-16}, and even this would mean that the mass of an electron would be doubled in the course of some 300,000,000, years, involving a slow but steady increase in the value of the gravitational constant.

The forcive TP may be decomposed into two parts T_1, P_1, due to electrons outside the ring, and T_2, P_2, due to the remaining electrons of the ring. T_2, P_2 are then expressed in terms of convergent Bessel series; and it is pointed out that in the case of steady motion T_1 is null, since it consists partly of a steady electric forcive, which vanishes because the lines of force of a steady electric field cannot form closed curves, and partly of a steady magnetic forcive, which vanishes because the mechanical action of a magnetic field on a moving electron is always perpendicular to its velocity. The resulting equations of motion are

$$nU = -\frac{r^2}{e^2}\frac{d(mv)}{dt} \quad . \quad . \quad . \quad . \quad (1)$$

$$P_1 - \frac{e^2}{r^2}(1+\beta^2)S + \frac{e^2}{r^2}V = \frac{mv^2}{r} \quad . \quad . \quad . \quad (2)$$

where

$$S = \frac{1}{4}\sum_1^{n-1} i \operatorname{cosec}\frac{\pi i}{n},$$

$$U = 2\sum_1^{\infty} s\left\{ sn\beta^2 J'_{2sn}(2sn\beta) - s^2n^2(1-\beta^2)\int_0^\beta J_{2sn}(2snx)dx \right\},$$

and V is another Bessel series, both U and V being convergent for β less than 1. For large values of n and small values of β, U converges so rapidly that it is sufficient to take the first term only.

For given values of n, β, r, (1) determines $\dot{\beta}$, and (2) determines r when β is given, but leaves β itself arbitrary. There is therefore nothing to fix the constitution of the atom, supposed to consist of invariable electrons in orbital motion with only electromagnetic forces between them. If, however, the electron varies slowly in radius, then

$$\frac{dm}{dt} = -m\frac{\dot{a}}{a} = -\frac{e^2}{c^2}\psi(\beta)\frac{\dot{a}}{a^2},$$

so that (1) becomes

$$\frac{nU}{\beta} = \frac{r^2\psi(\beta)}{c}\,\frac{\dot{a}}{a^2} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

When n and $\dot{a}/a^2$ are given, β and r are determined by (2) and (3). Since U is essentially positive, the average value of $\dot{a}$ must be positive. The numerical value of $\dot{a}/a$ may, however, vary considerably without sensibly affecting β when n is at all large, as is shown by the calculated values in the accompanying table:—

n		10	20	50	100	1000
β	$\frac{\dot{a}}{a}=10^{-16}$	·023	·131	·398	·566	·894
	$\frac{\dot{a}}{a}=10^{-46}$	·0074	·023	·187	·398	·833

We now have to consider the manner in which the mechanical relations between the electron and the surrounding ether must be modified in order to allow for the expansion of the electron. When the latter is invariable Dr Schott writes the energy equation in the form

$$v\,.\,\mathrm{T}=\frac{d\mathrm{E}}{dt}+\mathrm{R},$$

where $v\,.\,\mathrm{T}$ represents the rate of working of the mechanical forces on the electron; E is a function of its velocity, mass, and acceleration; and R is the rate at which the electron loses energy by radiation. E includes the electromagnetic energy of the electron, and $d\mathrm{E}$ is a perfect differential, while $\mathrm{R}dt$ is not. In the case of an expanding or contracting electron the equation must be replaced by

$$v\,.\,\mathrm{T}=\frac{d\mathrm{E}}{dt}+\mathrm{R}+\frac{e^2\dot{a}}{a^2}f(\beta),$$

where $f(\beta)$, like $\psi(\beta)$, depends on the structure of the electron. This additional term represents the rate at which work is done by, or against, the internal electromotive forces of the electron on account of the change of radius, and the existence in the electron of internal stresses resisting the change of volume must be assumed. Dr Schott finds, on calculating the internal E.M.F., that the principal term is derived from a potential Φ, the *convection potential* of Searle and Lorentz, which has the same value as if the electron were moving uniformly, and without relative changes in its parts, with its actual velocity. The remaining terms in the expression representing the E.M.F., due

to acceleration and to relative motion, are of the order a/ρ only, relatively to the principal term, and may be neglected. If ρ' be the electric volume density, then the stress-system P, Q, R, S, T, U, which balances the internal E.M.F., will therefore be determined by the equations

$$\frac{\partial P}{\partial x}+\frac{\partial U}{\partial y}+\frac{\partial T}{\partial z}=\rho'\frac{\partial \Phi}{\partial x},$$

$$\frac{\partial U}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial S}{\partial z}=\rho'\frac{\partial \Phi}{\partial y},$$

$$\frac{\partial T}{\partial x}+\frac{\partial S}{\partial y}+\frac{\partial R}{\partial z}=\rho'\frac{\partial \Phi}{\partial z}.$$

If ρ' be constant, or a funtion of Φ only, this stress-system reduces to a hydrostatic pressure

$$p=-\int\rho' d\Phi.$$

It would be possible to account for the above stress-system by assuming it to arise from attractions between elements constituting the electron, and this corresponds to the assumption of additional forces between electric charges made by Jeans[1] in order to account for the definite scale of atomic structure. These forces would, however, themselves require accounting for, and, like the additional forces assumed by Zöllner and Lorentz to account for gravitation (p. 423), would involve the abandonment of the definiteness and simplicity of the electron theory; and those now under consideration could not follow the law of inverse square, as otherwise they would balance the electrostatic repulsion between two similar electrons at rest. They would, therefore, not be capable of accounting for gravitation.

It is, therefore, preferable to assume the stress to arise from interaction between the electron and the ether; and the simplest assumptions will be, that the stress is a hydrostatic pressure, and that the surface of the electron is a surface of constant pressure. The electron will then be spherical when at rest, and when in motion, by the law of Lorentz and Fitzgerald, will be transformed into an oblate spheroid of eccentricity β (p. 445), having its least axis in the line of motion. Then, although, as we have seen, the ether will resist the expansion, or assist the contraction, of the electron with a pressure depending on its size and its velocity of translation, yet the resultant action of the ether on the electron vanishes, so that its motion through the ether is unaffected by the expansion or contraction. From our

[1] *Phil. Mag.*, vol. ii., 1901, p. 421; and vol. xi., 1906, p. 604.

present point of view the nucleus of the electron may either be a vacuole in the ether or may be filled with ether. The latter assumption necessitates some compressibility and expansibility in the ether, but the bulk modulus must be great enough for the inequalities of pressure due to expansion to be carried away by waves almost instantaneously. If this were not the case, the expanding or contracting electron would give rise to an increased or decreased pressure at a distant point which would vary, approximately, inversely as the distance from the electron, and would therefore give rise to a repulsion between two different electrons, varying inversely as the square of the distance between them.

In the rotationally elastic M'Cullagh ether the only permanent motions are irrotational, and the expanding or contracting electron will therefore constitute a hydrodynamic source or sink in the ether, supposed, as before, to be at rest at infinity. The velocity potential due to expansion at the rate $\dot{a}$ will be $\phi = -a^2\dot{a}/r$, so that

$$\sigma = -4\pi r\phi_1 = 4\pi a^2\dot{a}.$$

The gravitational attraction between two electrons $a_1\sigma_1\phi_1$ and $a_2\sigma_2\phi_2$ at distance r will therefore be

$$\frac{\rho\sigma_1\sigma_2}{4\pi r^2} = \frac{4\pi\rho a_1{}^2\dot{a}_1 a_2{}^2\dot{a}_2}{r^2} \quad . \quad . \quad . \quad (4)$$

Now consider two masses M_1 and M_2, the mass-centres of which are separated by a distance r, and which contain, respectively, n_1 electrons each of mass m_1, and n_2 electrons each of mass m_2. If $\dot{a}_1$ and $\dot{a}_2$ are both positive or both negative, the hydrodynamical attraction between them will be

$$\frac{M_1 M_2}{r^2} \cdot \frac{4\pi\rho a_1{}^2\dot{a}_1 a_2{}^2\dot{a}_2}{m_1 m_2},$$

which will correspond with the observed gravitational attraction if the second fraction have the value $G = 6{\cdot}7 \times 10^{-8}$.

If $\dot{a}_1$ and $\dot{a}_2$ are of opposite sign the force will be repulsive. If the masses are in motion, the values of m_1 and m_2, as also of $\dot{a}_1$ and $\dot{a}_2$, will depend on the velocity. The variation is, however, very small unless β approaches unity, which would only be the case for electrons grouped into rings each containing a very large number of electrons. As β approaches unity the gravitational attraction will diminish, because the pressure difference at the surface of the electron decreases and the mass increases. Rings consisting of few electrons, which move slowly, therefore produce practically the same effect as if the electrons were at rest; and

rings consisting of many electrons, which move quickly, produce a smaller effect. The difference is, however, very slight for a considerable difference in the number of electrons in the ring. Since an atom, if it consist, as we have assumed, of electrons in orbital motion, probably contains electrons moving with very different velocities, the gravitational attraction between two atoms will have an average value taken for these different velocities. It is not inconceivable that this average should be very nearly the same for atoms containing very different numbers of electrons, and therefore having very different atomic weights, and so make it possible to account for the observed constancy of gravitational action for different kinds of matter. This constancy would, however, be much more satisfactorily accounted for by Sir Oliver Lodge's suggestion (p. 407) that the main bulk of the atom may very possibly consist of a close admixture of positive electrons forming

"A continuous mass in the midst of which one or more isolated and individualised electrons may move about and carry on that display of external activity which confers upon the atom its observed properties."

With reference to his own theory, Dr Schott observes:

"While we may admit that the hypothesis of the expanding electron affords a qualitative explanation of the law of gravitation, yet we can hardly rest satisfied unless it gives for the gravitation constant a value of the right order. After many trials I have not succeeded in devising a mechanical model of the electron which shall do this in a satisfactory manner. When we treat the problem as that of a source in an incompressible liquid with the pressure at infinity equal to a given constant, we find that a is an elliptic function of the time, which varies between two definite limits. Since $\dot{a}/a$ must be small, a must be very close to one of the two limits; thus there are possible two electrons of different sizes, each varying very slowly in size; their radii being different, so also are their electromagnetic masses; by identifying them with the negative and positive electrons we can account for the difference in mass of the two kinds of electrons. In order to account for the smallness of $\dot{a}/a$ we must suppose the density ρ of the ether to be very large."

Dr Schott obtains this result as follows[1]:—

Let Π be the constant ether pressure at infinity, and suppose for simplicity that the nucleus of the electron consists of a vacuole in the ether. Then the pressure equation is

$$\frac{p}{\rho}+\frac{1}{2}\left(\frac{\partial\phi}{\partial r}\right)^2+\frac{\partial\phi}{\partial t}=\frac{\Pi}{\rho},$$

where the values of p, ϕ and $\partial\phi/\partial r$ are taken for $r=a$. This gives

$$\frac{p}{\rho}=-a\ddot{a}-\frac{3}{2}\dot{a}^2=\frac{\Pi}{\rho}.$$

[1] Communicated to the author in a letter from Dr Schott.

Now p must be the Maxwell stress due to the charge e, so that $p = e^2/8\pi a^4$, and therefore

$$\frac{e^2}{8\pi\rho a^4} - a\ddot{a} - \frac{3}{2}\dot{a}^2 = \frac{\Pi}{\rho},$$

and therefore

$$\frac{e^2}{4\pi\rho}\frac{\dot{a}}{a^2} - 2a^3\dot{a}\ddot{a} - 3a^2\dot{a}^3 = \frac{2\Pi}{\rho}a^2\dot{a}.$$

Hence, integrating,

$$C - \frac{e^2}{4\pi\rho a} - a^3\dot{a}^2 = \frac{2\Pi}{3\rho}a^3,$$

or

$$a^4\dot{a}^2 = Ca - \frac{e^2}{4\pi\rho} - \frac{2\Pi}{3\rho}a^4,$$

and therefore

$$t = \pm\int\frac{a^2 da}{\sqrt{Ca - \frac{e^2}{4\pi\rho} - \frac{2\Pi}{3\rho}a^4}}$$

which gives a as an elliptic function of t. It is easily seen that the expression under the square root vanishes only for two real positive values of a, say a_1 and a_2, between which a can oscillate backwards and forwards. The period of the oscillation would have to be very large, and the present state of the universe would, on this view, be merely a phase in the oscillation, in which the negative electrons were expanding and the positive electrons contracting, if the whole volume of the ether is to remain constant. The end of the present phase would be reached when the negative electrons reached their upper limit and the positive electrons reached their lower limit. It is not difficult to choose the constants C and Π so as to make one of these limits very much larger than the other. The larger might then correspond to a positive electron and the smaller to a negative. Taking $\dot{a}/a = 10^{-10}$ would give a variation of about $\frac{1}{3}$ per cent. per annum, and would require a value of ρ of the order 10^{100}. The variation of $\dot{a}/a$ might be diminished as much as we pleased by assuming ρ to be sufficiently great. In itself there does not appear to be any objection to such a large value of ρ, but it would be an additional assumption. The value 10^{12} obtained by Sir Oliver Lodge (p. 184) was, as we saw, merely a minimum; there is every reason to suppose it greater, and I know of no valid argument against its being very considerably greater. Dr Schott says that

"It leads to a difficulty, which however is not peculiar to the present view, but is shared by every theory which assumes a mechanical action between the electron and the ether. A sphere moving through a liquid

has an effective mass greater than its true mass by one half the mass of liquid displaced. If this were true of the electron, its effective mass would be enormously greater than its electromagnetic mass. For the negative electron this is not the case; it moves through the ether without experiencing any appreciable resistance beyond that due to electromagnetic radiation."

This would be the natural inference to be drawn if Lorentz's purely mathematical concept of the electron as a sphere carrying a charge of electricity were regarded as physically representing an electron. That it is not even a possible physical representation I pointed out on page 182; and the comparison of the results of Kaufmann's experiments with Lodge's calculations of the mass of a spherical electron (p. 184), showing that the mass determined from the magnetic force which it establishes and carries with it in its motion is twice the mass of the ether contained in the sphere as compared with one and a half times the mass determined hydrodynamically, appears to me to prove conclusively that the hydrodynamic mass of the negative electron is negligible in comparison with the electromagnetic mass, and that this difficulty arises only from pressing too far the mathematical analogy of the charged sphere. Whether or not this hold good for the positive electron also, the calculations referred to show that the hydrodynamic and electromagnetic mass would be affected in the same ratio by any change in the density of the ether. A much more formidable objection, in my opinion, to the hypothesis of a slow periodic expansion and contraction of the electrons is that the equations of motion, (1) and (2), can be satisfied only during the expansion, and not during the contraction, of the electrons. For we saw that the series U is essentially positive, and the electromagnetic mass is also essentially positive, since it is proportional to the square of the charge. Dr Schott is therefore obliged to assume that the phase of expansion of the negative electron would be a phase of stability of rings of electrons, the phase of contraction one of instability, the first a period of building up of matter, the second a period of breaking down. Dr Schott does not consider such a conclusion inadmissible, and if the period available could be made sufficiently long it might not be so; but in my opinion it would form a strong argument against the hypothesis of expanding electrons, even as a provisional hypothesis, if any equally simple hypothesis, not involving such a conclusion, can be shown to offer a possible explanation of the observed facts. Moreover, in order to account for gravitation on this hypothesis, an enormously high value would have to be assumed for the density of the ether. This objection might be met by assuming pulsations to be superposed upon the expanding electrons in order

to account for gravitation, but this would mean the adoption of two distinct hypotheses, one to account for gravitation, and the other for the stability of matter, neither of which was required for any other purpose. Another difficulty which I had pointed out to Dr Schott, before learning from him that he could see no way out of the difficulty referred to above, was that even if both the determination of the atomic scale and gravitation could be accounted for by alternate phases of slow expansion and contraction of the electrons, gravitation, and therefore cohesion, would, at each turning-point, necessarily fall, for a finite time, to a value sensibly equal to zero, which would involve the disintegration of all the matter in the universe.

This difficulty will be overcome if gravitation be ascribed, as in the text, to the hydrodynamic, instead of the electromagnetic, mass of the electrons, since the hydrodynamic mass, unlike the electromagnetic, changes sign when the expansion becomes a contraction. The objection of the temporary vanishing of gravitation and cohesion at the turning-point would, however, still remain. Dr Schott raises the further objection that the assumption that the electron has hydrodynamic as well as electromagnetic mass would lead to a conflict with the experimental determinations of Kaufmann and others of the speeds and masses of negative electrons. In this, however, I cannot agree with him. The extremely small value of the gravitational, as compared with the electric, force between two electrons (p. 419), shows that a hydrodynamic mass, far too small to be detected in these experiments, would be sufficient to account for gravitation, on the basis of Sir Oliver Lodge's conclusion (p. 185) that the electromagnetic, like the hydrodynamic, mass is proportional to the density of the ether. Further, the identity, as regards dimensions, of the electromagnetic and hydrodynamic mass there arrived at, combined with the variation in the latter with the motion of the electron, leads to the conclusion that equations of motion similar to those obtained by Dr Schott for expanding electrons would apply to electrons forming permanent ether sources or sinks, so that the latter hypothesis would account both for gravitation and the stability of matter, with the additional advantage of providing for the existence of negative, as well as positive, gravitating matter.

We now have to consider the theory of rapid pulsations, due to Bjerknes and Hicks. Before attacking the problem of the mutual action of two pulsating spheres in an infinite incompressible fluid, Professor Hicks had shown[1] that if a source of strength σ

[1] *Roy. Soc. Proc.*, 1897, and *Phil. Trans.*, 1880; or Basset's *Hydrodynamics*, vol. i. p. 51.

exist at a point P in such a fluid, then the image of the source with respect to a sphere of radius a with its centre at a point O consists of a source of strength $\sigma a/\mathrm{OP}$ at the inverse point of P, and which therefore coincides with the electrical image of P with respect to the sphere, and a line sink of strength σ/a per unit of length, extending from the inverse point of P to the centre of the sphere. If then there are two spheres A_1 and A_2 in the fluid, of which A_1 is pulsating, then, at any time, the motion of the fluid due to A_1 alone would be that of a source at the centre of A_1 of strength $\sigma = -a^2\dot{a}$. The simultaneous action of pulsations of A_2 may then be represented by the addition of an infinite series of images; the first being the image of the source representing the action of A_1 with respect to the sphere A_2, as defined above, within A_2; then an image within A_1 of the first image within A_2; and so on indefinitely. Hicks has found [1] a remarkable relation between the successive images which makes it possible to determine them all. The actions of A_1 on A_2 and of A_2 on A_1 can then be expressed, by the aid of formulæ established in the papers previously referred to, in terms of the distance of the spheres and of their radii and their time rates.[2] Let a_1 and a_2 be the radii of the two spheres, r the distance between their centres, p_1 the resultant pressure of the fluid on A_1 towards A_2, p_2 the resultant pressure of the fluid on A_2 towards A_1, m_1 the mass of fluid displaced by A_1, and m_2 the mass displaced by A_2. Then, when the calculation of portions of the pressure depending on the square of the velocity of the fluid is carried as far as the term r^{-5}, it is found that, where ρ is the density of the fluid,

$$p_1 = -2\pi a_1{}^2\frac{d}{dt}(\mathrm{M}_1+\mathrm{N}_1) - \frac{\rho a_2{}^3 a_1{}^4 \dot{a}_1{}^2}{r^5},$$

$$p_2 = -2\pi a_2{}^2\frac{d}{dt}(\mathrm{M}_2+\mathrm{N}_2) - \frac{\rho a_1{}^3 a_2{}^4 \dot{a}_2{}^2}{r^5},$$

where

$$\mathrm{M}_2 = \frac{\dot{m}_1 a_2}{4\pi r^2}\left\{1 + \frac{a_1{}^3 a_2{}^3}{(r^2-a_1{}^2)(r^2-a_1{}^2-a_2{}^2)^2} + \frac{a_1{}^6 a_2{}^6}{\{(r^2-a_1{}^2)^2-a_2{}^2r^2\}\{(r^2-a_2{}^2)^2-2a_1{}^2r^2+a_1{}^2(a_1{}^2+a_2{}^2)^2\}}+\ldots\right\},$$

$$\mathrm{N}_2 = \frac{\dot{m}_2 a_1{}^3 a_2}{4\pi r}\left\{\frac{1}{(r^2-a_1{}^2)^2} + \frac{a_1{}^3 a_2{}^3}{(r^2-a_1{}^2-a_2{}^2)\{(r^2-a_1{}^2)^2-a_2{}^2r^2\}^2}+\ldots\right\},$$

[1] *Camb. Phil. Soc. Proc.*, vol. iii. p. 277, 1880.

[2] See also Basset's *Hydrodynamics*, vol. i. p. 248.

and M_1, N_1 are obtained from M_2, N_2 by symmetrically interchanging a_1 and a_2. If we neglect all powers of $1/r$ above the second we shall have

$$\frac{p_2}{\rho} = -\frac{2\pi a_2^2}{r^2}\frac{d}{dt}(a_1^2 a_2 \dot{a}_1).$$

Let

$$a_1 = \bar{a}_1 + a \sin\frac{2\pi t}{\tau},$$

$$a_2 = \bar{a}_2 + \beta \sin\frac{2\pi}{\tau}(t-\epsilon),$$

so that $\bar{a}_1$ and $\bar{a}_2$ are the mean values of the radii. Then, if $\bar{p}_1$ and $\bar{p}_2$ be the mean values of p_1 and p_2, we shall have

$$\begin{aligned}\frac{\bar{p}_2}{\rho} = -\frac{2\pi}{r^2\tau}\int_0^\tau a_2^2\frac{d}{dt}(a_1^2 a_2\dot{a}_1)dt = \frac{4\pi}{r^2\tau}\int_0^\tau a_1^2 a_2^2 \dot{a}_1\dot{a}_2 dt \\ = \frac{16\pi^3}{r^2\tau^3}\bar{a}_1^2\bar{a}_2^2 a\beta\int_0^\tau \cos\frac{2\pi t}{\tau}\cos\frac{2\pi}{\tau}(t-\epsilon)dt \\ = \frac{8\pi^3\bar{a}_1^2\bar{a}_2^2 a\beta}{r^2\tau^2}\cos\frac{2\pi\epsilon}{\tau} = \frac{\bar{p}_1}{\rho}. \qquad (5)\end{aligned}$$

It follows, therefore, that if the spheres are pulsating with equal periods, they will attract each other when their phases differ by less than a quarter of a period; but if the phases differ by more than a quarter and less than three-quarters of a period, they will repel each other.

Professor Hicks points out that the property possessed by two pulsating spheres in a fluid of acting on each other with a force whose principal part varies inversely as the square of the distance, belongs to all pulsating bodies. This follows from the fact, which was first pointed out by Sir George Stokes,[1] that whenever there is a change of volume of a body immersed in a fluid, the expression for the velocity potential contains a term of the form A/r.

If we compare (5) with (4) we see that rapid pulsations of electrons whose distances are great in comparison with their radii, and we saw on p. 196 that this holds good even with respect to the electrons in the same atom, give rise to an instantaneous distribution of pressure identical with that due to slowly expanding electrons forming practically steady hydrodynamic sources.

[1] "On Some Cases of Fluid Motion," *Camb. Phil. Trans.*, 1849.

We saw, moreover, that the equations of motion for permanent sources or sinks are of the same form as for expanding electrons. It therefore follows that similar equations will be obtained in the case of pulsating electrons in an atom, provided only the pressure is, at every instant, sensibly the same on all the electrons within any one atom; and the fulfilment of this condition is a necessary result of the initial assumption that the compressibility of the ether, if of finite amount, is exceedingly small.

APPENDIX Q.

HAECKEL'S "RIDDLE OF THE UNIVERSE."

THE SO-CALLED LAW OF SUBSTANCE (Chapter XII.).

"THE *law of the persistence* or *indestructibility of matter*, established by Lavoisier in 1789, may be formulated thus: The sum of matter, which fills infinite space, is unchangeable. . . .We may formulate the *law of the persistence of force* or *conservation of energy* thus: The sum of force, which is at work in infinite space and produces all phenomena, is unchangeable" (p. 208). "Modern physics draws a distinction between 'force' and 'energy,' but our general observations so far have not needed a reference to it" (p. 218). "The unity of the two laws—still much disputed—is expressed by many scientists who are convinced of it in the formula: 'Law of the persistence of matter and force.' . . . In the ultimate analysis it is found to be a necessary consequence of the principle of causality" (p. 219).

The assumption that the absolutely meaningless expression persistence of force is equivalent to the principle of conservation of energy shows that Professor Haeckel does not bring to his task even the elementary knowledge of physics which a first-year student is expected to possess, so that we are not surprised to find him (p. 220) defining energy as moving force. The inherent absurdity of the last sentence quoted makes it perfectly clear that Professor Haeckel is as weak in philosophy as in physics. His ideas of the ether, taken from an unscientific German writer, are set forth (p. 222) as follows:—

"In fundamental opposition to the theory of vibration, or the kinetic theory of substance, we have the modern 'theory of condensation,' or the pyknotic theory of substance. It is most ably established by J. G. Vogt,[1] in *The Nature of Electricity and Magnetism on the Basis of a Simplified Conception of Substance* (1891). Vogt assumes the primitive force of the world . . . to be . . . the condensation of a simple primitive substance, which fills the infinity of space in an unbroken continuity. Its sole inherent mechanical form of activity consists in a tendency to condensa-

[1] This writer's work, as was pointed out on p. 504, is founded on the entirely unscientific pre-Newtonian ideas of the ether.

tion or contraction, which produces infinitesimal centres of condensation. . . . These . . . . correspond in general to the ultimate separate atoms of the kinetic theory; they differ, however, very considerably in that they are credited with sensation and inclination (or will-movement of the simplest form), *with* souls in a certain sense, in harmony with the old theory of Empedocles of the 'love and hatred of the elements.' Moreover, these 'atoms with souls' do not float in empty space, but in the continuous, extremely attentuated intermediate substance which represents the uncondensed portion of the primitive matter. By means of certain 'constellations, centres of perturbation, or systems of deformation,' great masses of centres of condensation quickly unite in immense proportions, and so obtain a preponderance over the surrounding masses. By that process, the primitive substance, which in its original state of quiescence had the same mean consistency throughout, divides or differentiates into two kinds. The centres of disturbance, which *positively* exceed the mean consistency in virtue of the *pyknosis* or condensation, form the ponderable matter of bodies; the finer, intermediate substance, which occupies the space between them, and *negatively* falls below the mean consistency, forms the ether, or imponderable matter. As a consequence of the division into mass and ether there ensues a ceaseless struggle between the two antagonistic elements, and this struggle is the source of all physical processes. The positive ponderable matter, the element with the feeling of like or desire, is continually striving to complete the process of condensation, and thus collecting an enormous amount of *potential* energy; the negative, imponderable matter, on the other hand, offers a perpetual and equal resistance to the further increase of its strain and of the feeling of dislike connected therewith, and thus gathers the utmost amount of actual energy" (p. 223). "As to my own opinion—and that of many other scientists—I must lay down the following theses, which are involved in Vogt's pyknotic theory, as indispensable for a truly monistic view of substance, and one that covers the whole field of organic and inorganic nature:—I. The two fundamental forms of substance, ponderable matter and ether, are not dead and only moved by extrinsic force, but they are endowed with sensation and will[1] (though, naturally, of the lowest grade): they experience an inclination for condensation, a dislike of strain; they strive after the one and struggle against the other" (p. 224).

Such writing would require the caustic wit of De Morgan to do it justice, and it would be amongst congenial surroundings in his *Budget of Paradoxes*. From all its obscurity, however, two facts must be patent to every reader. The statement that matter is not moved only by external force is in direct contradiction with each of Newton's three laws of motion, and is therefore incompatible with the observational basis of dynamics; and a hypothesis

[1] Professor Haeckel informs us on another page that he regards this sensation and will as unconscious; but this makes the terms meaningless, for it is as impossible to ascribe meanings to the word-combinations *unconscious sensation* and *unconscious will* as to the combinations *square circle* and *triangular square*.

which ascribes mental powers to the ultimate atoms, in such flagrant disregard of experience, affords a striking exemplification of the statement made on p. 461, that the attempt to ignore the mental scheme in the development of any scheme attempting to account for the order of Nature will invariably be found to necessitate its introduction in some disguised and unscientific manner.

The Law of Entropy or Dissipation of Energy (Chapter XIII.).

"This infinite and eternal machine of the universe sustains itself in eternal and uninterrupted movement, because every impediment is compensated by an 'equivalence of energy' and the unlimited sum of kinetic and potential energy remains always the same. The law of the persistence of force proves also that the idea of a *perpetuum mobile* is just as applicable to, and as significant for, the cosmos as a whole as it is impossible for the isolated action of any part of it. Hence the theory of *entropy* is likewise untenable" (p. 252). "If this theory of entropy were true, we should have a 'beginning' corresponding to this assumed 'end' of the world—a minimum of entropy, in which the differences in temperature of the various parts of the cosmos would be at a maximum. Both ideas are quite untenable in the light of our monistic and consistent theory of the eternal cosmogenetic process; both contradict the law of substance. There is neither beginning nor end of the world" (p. 253).

Haeckel's treatment of the observed principle of the Dissipation of Energy affords another example of the remarkable contrast between his view of the value of hypotheses and the scientific point of view which I have endeavoured to make clear in this volume (see, for example, pp. 57 and 447). Haeckel does not bring forward the argument that the truth of the law of increasing entropy was only proved for the small portion of the universe accessible to observation, and that its validity for the whole universe was not a necessary consequence, although this argument would have been quite legitimate in the state of scientific knowledge in the year 1899, when the *Riddle of the Universe* was written. He simply rejects the observed law because it contradicts his hypotheses. I will give two further extracts in which he refers to purely hypothetical assumptions as deciding between two conflicting views of Nature. Such a view is too funny for words.

The Carbon Theory and the Origin of Life (Chapter XIV.).

"The peculiar chemico-physical properties of carbon—especially the fluidity and the facility of decomposition of the most elaborate albuminoid compounds of carbon—are the sole and the mechanical causes of the specific phenomena of movement which distinguish

organic from inorganic substances, and which are called life, in the usual sense of the word. Although the 'carbon theory' is warmly disputed in some quarters, no better monistic theory has yet appeared to explain it" (p. 262). "The hypothesis of spontaneous generation and the allied carbon theory are of great importance in deciding the long-standing conflict between the *teleological* (dualistic) and the *mechanical* (monistic) interpretation of phenomena" (p. 264).

Haeckel's idea of logical argument, as adapted for the readers to whom this work is addressed, is as strange as are his views on the object and utility of hypotheses. and equally so is his conception of the nature of logic. I will take the latter point first.

"What we call the soul is, in my opinion, a natural phenomenon; I therefore consider psychology to be a branch of natural science—a section of physiology" (Chapter VI., p. 91). "The first task of every science is the clear definition of the object it has to investigate. In no science, however, is this preliminary task so difficult as in psychology, and this circumstance is the more remarkable since logic, the science of defining, is itself a part of psychology."

As Professor Münsterberg, of Harvard, tells us (*Psychology and Life*, p. 37):

"We are told that we are to expect an exact knowledge of the psychical facts from our knowledge of the brain; but what in the world can we know better than the objects of our immediate self-observation? The observation and analysis of our mental facts is in no way dependent upon a hypothesis in regard to the soul; it is the most direct object of our attention, and we thus know endlessly more about our psychical facts than about the functions of the brain. Even two thousand years ago the chief mental facts were well known, while the most fundamental questions of brain physiology are still to-day under lively discussion. Above all, the history of science shows how in the times of their co-operation psychology always had to give and physiology to take; light had to be thrown from the side of the well-known psychological facts upon the obscure physiological facts, and never in the opposite direction. The consequence of this situation is that psychologists in their work of analysis and research into the constitution of the psychical facts have not the slightest reason for inquiring into any accompanying brain processes; they cannot learn from that side anything which they do not know better from self-observation and the observation of others."

The statement that psychology is a section of physiology shows, therefore, that Professor Haeckel is fundamentally ignorant of the meaning of psychology; but what are we to say of the conclusion which follows of necessity from his two statements taken together, viz., that logic is part of a section of physiology? Would it be within the bounds of possibility to make a statement embodying greater absurdity and topsyturvydom? As an example of what is presented as argument in this work, I will

take the method of dealing with the immortality of the soul, the possibility of which Professor Haeckel repeatedly asserts that he has scientifically disproved. He first asks, among other questions, What really is the soul? What is its relation to the mind? (Chapter VI., p. 94). He does not attempt, in this chapter, to answer his own question, but proceeds to supply his readers with a misrepresentation of Kant's position so gross that I can only suppose Professor Haeckel's knowledge of this great thinker to have been derived entirely from the study of some magazine article giving a travesty of his arguments. In Chapter VI. (p. 92) Professor Haeckel had promised to give his demonstration in Chapter XI. In the meantime (Chapter VII., p. 111) he defines the soul as an abstraction, so that the question of its immortality would be a meaningless one. In Chapter XI. he first assumes it to be a material substance, and then makes a series of assertions which he describes as a brief exposition of sound scientific arguments The utter ineptitude of the so-called argument will appear most clearly by reproducing his statements without interruption by further comments. The first paragraph is the reply to his own question on p. 94.

"The answers to these and many other cognate questions are infinitely varied: not only are the views of the most eminent thinkers on these questions widely divergent, but even the same scientific authority has often completely changed his views in the course of his psychological development. . . . The most interesting example of such an entire change of objective and subjective psychological opinions is found in the case of the most influential leader of German philosophy, Immanuel Kant. The young, severely *critical* Kant came to the conclusion that the three great buttresses of mysticism—'God, freedom, and immortality'—were untenable in the light of 'pure reason'; the older, *dogmatic* Kant found that these three great hallucinations were postulates of 'practical reason,' and were, as such, indispensable" (p. 94).

"All the phenomena of the psychic life are, without exception, bound up with certain material changes in the living substance of the body, the *protoplasm*. We have given to that part of the protoplasm which seems to be the indispensable substratum of psychic life the name of *psychoplasm* (the 'soul-substance' in the monistic sense); in other words, we do not attribute any peculiar 'essence' to it, but we consider the psyche to be merely a *collective idea of all the psychic functions of protoplasm*. In this sense the 'soul' is merely a physiological abstraction like 'assimilation' or 'generation'" (p. iii).

"The conception of the soul as a 'substance' is far from clear to many psychologists; sometimes it is regarded as an 'immaterial' entity of a peculiar character in an abstract and idealistic sense, and sometimes as a confused 'tertium quid' between the two. If we adhere to the monistic idea of substance . . . . we find energy and matter inseparably associated in it. We must, therefore, distinguish in the 'substance of the soul' the characteristic psychic *energy* which is all we perceive

(sensation, presentation, volition, etc.), and the psychic *matter* which is the indispensable basis of its activity, that is, the living protoplasm. Thus in the higher animals the 'matter' of the soul is a part of the nervous system; in the lower, nerveless animals and plants it is a part of their multicellular protoplasmic body; and in the unicellular protists it is a part of their protoplasmic cell-body. In this way we are brought once more to the psychic organs, and to an appreciation of the fact that these material organs are indispensable for the action of the soul; but the soul itself is *actual*—it is the sum-total of their physiological functions" (Chapter XI., p. 202).

"When we come to analyse all the different proofs that have been urged for the immortality of the soul, we find that not a single one of them is of a scientific character; not a single one is consistent with the truths we have learned in the last few decades from physiological psychology and the theory of descent. The *theological* proof—that a personal creator has breathed an immortal soul (generally regarded as a portion of the divine soul) into man—is a pure myth. The *cosmological* proof—that the 'moral order of the world' demands the eternal duration of the human soul—is a baseless dogma. The *teleological* proof—that the 'higher destiny' of man involves the perfecting of his defective, earthly soul beyond the grave—rests on a false anthropism. The *moral* proof—that the defects and the unsatisfied desires of earthly existence must be fulfilled by 'compensative justice' on the other side of eternity —is nothing more than a pious wish. The *enthnological* proof—that the belief in immortality, like the belief in God, is an innate truth, common to all humanity—is an error in fact. The *ontological* proof—that the soul, being a simple, immaterial, and indivisible entity, cannot be involved in the corruption of death—is based on an entirely erroneous view of the psychic phenomena; it is a spiritualistic fallacy. All these and similar 'proofs of athanatism' are in a parlous condition; they are definitely annulled by the scientific criticism of the last few decades.

"The extreme importance of the subject leads us to oppose to these untenable 'proofs of immortality' a brief exposition of the sound scientific arguments against it. The *physiological* argument shows that the human soul is not an independent, immaterial substance, but, like the soul of all the higher animals, merely a collective title for the sum-total of man's cerebral functions; and these are just as much determined by physical and chemical processes as any of the other vital functions, and just as amenable to the law of substance. The *histological* argument is based on the extremely complicated microscopic structure of the brain; it shows us the true 'elementary organs of the soul' in the ganglionic cells. The *experimental* argument proves that the various functions of the soul are bound up with certain special parts of the brain, and cannot be exercised unless these are in a normal condition; if the areas are destroyed, their function is extinguished; and this is especially applicable to the 'organs of thought,' the four central instruments of mental activity. The *pathological* argument is the complement of the physiological: when certain parts of the brain (the centres of speech, sight, hearing, etc.) are destroyed by sickness, their activity (speech, vision, hearing, etc.) disappears; in this way nature herself makes the decisive physiological experiment. The *ontogenetic* argument puts before us the facts of the development of the soul in the

individual: we see how the child-soul gradually unfolds its various powers; the youth presents them in full bloom, the mature man shows their ripe fruit; in old age we see the gradual decay of the psychic powers, corresponding to the senile degeneration of the brain. The *phylogenetic* argument derives its strength from palæontology and the comparative anatomy and physiology of the brain; co-operative with and completing each other, these sciences prove to the hilt that the human brain (and, consequently, its function—the soul) has been evolved step by step from that of the mammal, and, still further back, from that of the lower vertebrate.

"These inquiries, which might be supplemented by many other results of modern science, prove the old dogma of the immortality of the soul to be absolutely untenable."

The reader can see for himself that there is here not even the semblance of an argument, but merely a series of assertions, concluding with the assertion that the thesis is proved. These paragraphs are reproduced only as illustrating the kind of verbiage which Haeckel presents to his readers as argument. It will not, however, be amiss to compare with the assertions of the lesser biologist the words of a far greater one. Huxley tells us[1] that physical science "effectually closes the mouths of those who pretend to refute it [the continuance of the human spirit after death] by objections deduced from merely physical data." Huxley, in addition to being a great biologist, was a physicist and a philosopher, while biology is Haeckel's sole equipment.

In dealing with the great question of the liberty of the will, Haeckel does not profess to offer any argument, but contents himself with the assertions that the controversy has now ended completely in favour of the determinist, and that the liberty of the will has been disproved by comparative physiology and evolution. The first assertion is untrue, the controversy still rages as of old; the second assertion is obviously absurd. Readers who care for further serious criticism of the *Riddle of the Universe* may be referred to two excellent works:—Sir Oliver Lodge's little volume on *Life and Matter*; and the Rev. J. Gerard's *The Old Riddle and the Newest Answer*, which is obtainable in a cheap sixpenny edition. I may, however, conclude by pointing out the twofold nature of Haeckel's failure. In the first place, in common with many other able scientific craftsmen, he has failed to grasp the fact that all the quantitative concepts of physical science are simplifications in which the phenomena as presented to our consciousness are artificially transformed into elements capable of logical development into orderly working models, and that even if these conceptual models could be so perfected as to afford a representation, adequate to our intelligence and powers of observa-

[1] *Science and Morals*, p. 143.

tion, of all observed phenomena, they could not possibly provide a basis for denying the existence of realms beyond their limits. In the second place, in order to provide a wider basis for the physiological models which he is capable of dealing with scientifically, he attempts to introduce into the models forming the representations of the phenomena of molar and molecular physics, elements which would throw the whole models out of gear, simply because he is there dealing with subjects of which his knowledge is insufficient to enable him either to form an opinion for himself or to distinguish between competent and incompetent guides.

From one point of view the reader will find the serious consideration of the *Riddle of the Universe* by no means uninstructive, and that is an endeavour to discover the source of the worst inconsistencies, and an inquiry into the modifications which must be made in Haeckel's hypotheses in order to bring them into accordance with observation. This source is evidently the self-contradictory assumption of atoms possessing *unconscious will* and *unconscious sensation*. If a physicist were to amuse himself by building up a nonsense theory of the universe on the initial assumption that the *ultimate atoms are extremely minute, hard, spherical, soft cubes of immense size*, he would have plenty of properties to deal with; and assuming his readers to be sufficiently unintelligent to take his initial assumption seriously, he would merely have to concern himself with keeping mutually contradictory elements as far apart as possible. Where their juxtaposition was unavoidable he would make use of long, involved sentences, translating a fair proportion of the terms into some language not likely to be very familiar to the bulk of his readers. Professor Haeckel generally favours Greek for this purpose, and the mutual inconsistency of his initial assumptions forces him to adopt such artifices in order to construct something which looks like an argument. He has no intention of leading his readers to untrue conclusions, for it is perfectly clear that he himself is convinced of their truth. He is, however, most certainly not convinced by his own arguments, but by other quite inscrutable means. He accordingly endeavours, and apparently with much success, to drive his readers, by whatever can be made to look like an argument, towards the conclusions which he is perfectly satisfied in his own mind are the right ones. Now since Haeckel's scheme falls to pieces entirely if sensation and will be abandoned, we must retain them, and abandon the contradictory qualification *unconscious* and replace it by *conscious*. This clears the initial self-contradiction out of the scheme, and consciousness and intelligence of a much higher grade than human consciousness and intelligence are requisite to its success. For the atoms have a very difficult task to perform.

It would be quite difficult enough if they were only required to obey the observed laws of dynamics, for this would necessitate the exercise, on the part of every atom, of such will-power as to behave exactly as if it possessed neither will nor sensation until it happens to find itself part of some living body, when it may begin to exert its will, in due co-operation with the other atoms constituting the living body, so as to ensure to it such welfare as may be appropriate to its circumstances. Haeckel's scheme, however, demands a much higher grade of atomic intelligence, for the atoms also have to provide for the eternal duration of the universe by counteracting the dissipation of energy. Now the dissipation of energy is observed to hold good at the present time in our portion of the universe. The transformation of organised into unorganised energy in our portion of the universe must therefore be balanced by an exactly equal transformation of unorganised into organised energy elsewhere. The atoms responsible for this balancing must therefore possess a collective intelligence of the order which we found must be attributed, as a minimum, to the universal mind. According to Haeckel, every atom is to have the same intelligence, and therefore every atom in the universe must be credited with the possession of the minimum power which had to be attributed to the universal mind. We can overcome this difficulty by availing ourselves of the knowledge that all atoms are interconnected through the ether, which leads us to the simpler concept of a single intelligence possessing these powers. This brings us to the theory which I have set forth in Chapter XXIV. That is to say, the minimum alteration in Professor Haeckel's scheme which is necessary to bring it into logical consistency with itself and into agreement with observation transforms it into the theory there presented.

APPENDIX R.

A SIMPLE EXPERIMENT IN THOUGHT TRANSFERENCE, AND AN HYPOTHESIS BY WHICH IT MAY BE ACCOUNTED FOR.

THE experiment was made at home, the only persons present, and who all took part in it, being my wife, her sister, and one of my daughters. A statement is made in Hudson's *Law of Psychic Phenomena*, that amongst any half-dozen persons there is quite likely to be found one capable of telling the name of a card selected by the others. The procedure suggested is that after a card has been selected in the absence of the proposed thought-reader, all should join hands in a ring, and all except the latter, who of course does not know the card selected, should form as strong a mental picture of the card as they are capable of, and endeavour to keep their attention steadily upon it. The statement had been mentioned as a very remarkable one if there were any truth in it, and my daughter was anxious to try the experiment. I agreed to the request that I should attempt to name a card in this manner which the others were to select in my absence, although I had no expectation of a successful result. I left the room while the card was selected, then took my seat in an arm-chair, joined in the circle, closed my eyes to prevent distraction from external objects, and concentrated my mind upon the idea of receiving the impression. After some little time, the length of which was not noted, I seemed to see, presented to my mind, a very indistinct impression of a blank surface with a darker, but colourless, central blotch. Neither the surface nor that of the blotch presented any distinct outline, but from the size of the blotch I concluded that, if it meant anything, it must be the ace of spades, and asked doubtingly if that were the one selected. The selected card was the ace of hearts.

Now the one feature in the impression was the single central blotch, and it therefore seemed to me that the failure might not impossibly indicate a partial success, and it was therefore

determined to repeat the experiment, and this time we all entertained the idea of the definite possibility of a successful issue. The procedure was as before, except that I now decided not to attempt to form a definite idea of receiving a mental impression of a card, but simply to exclude, by concentration of the will, all ordinary sense perceptions, and also to inhibit all thoughts arising in the mind. As soon as this was completely effected I began to see faint but distinctly recognisable pictures of four different cards which appeared like badly focussed dissolving lantern pictures flitting into and out of vision. These were the jack, nine, five, and three of diamonds. As I required one, and not four, cards I waited patiently, without allowing these passing pictures to cause distraction or give rise to reflection. Presently I saw a perfectly distinct picture, like a well-focussed lantern slide, of three diamonds on a clearly outlined card; but the diamonds, instead of being in a vertical line, formed the vertices of an equilateral triangle. This picture again was rejected, as not representing any known card, and was instantly replaced, as one well-focussed lantern slide replaces another, by an equally distinct picture of the three of diamonds, which proved to be the selected card. Although the outlines of both cards and pips, and the markings on the jack, were perfectly readable throughout the second experiment, there was no sign of colour, the only impression retained being that of a monochrome picture.

The explanation of the intrusion of the three other diamonds mentioned above is not the least interesting feature of the experiment. Each of the selectors had originally proposed one of these cards, and it was finally decided amongst them to reject the three first selections, and agree upon a fourth, and it became perfectly evident that their intrusion was due to each of the selectors allowing her attention to oscillate between her original and her final choice. It is noteworthy, but not a fact of which any explanation occurs to me, that as long as the four pictures continued their alternation there was no perceptible distinction in definition in favour of the three of diamonds, and that as soon as the three became clearly defined the intrusions of the others ceased entirely. When I mentioned the first appearance of the three diamonds as a triangle, my sister-in-law stated that, finding herself incapable of keeping a clear picture of the card in her mind, she had concentrated her thought upon a diamond pendant with three diamonds in a triangle. Even the comparatively indistinct pictures seen when the four diamonds were presenting themselves in irregular succession resembled definite visual images totally different from pictures recalled by memory.

These details appear to me to eliminate completely the possibility

of the result being simply a coincidence. The mere selection of one card correctly out of the fifty-two would only have odds of fifty-one to one against it; but, even if I had known beforehand, which I did not, that there were four cards in question, and one to be chosen out specially amongst them, the odds against guessing the four correctly and correctly selecting one out of them again would have been about twenty-five millions to one. Moreover, the hypothesis of a lucky guess would not account for the visual impressions observed.

I endeavoured to follow this experiment by another in which an object (a ship) was thought of, and hands were not joined. This was entirely unsuccessful, but the want of success could not be taken as affording appreciable support to the hypothesis, which seems to be very generally accepted, that the thought-transference is facilitated by bodily contact, as the want of success may have been due to quite other causes.

The absence of colour in the visual pictures appears to be a point worthy of attention, in view of experiments made by Captain Abney,[1] which led him to conclude that the stimulus required to produce colour is of a different order from that required to produce light.

The close resemblance of the mental impressions to visual pictures suggests their being due to some kind of radiation emitted from the brains of those who were concentrating their attention upon the image which they desired to transmit. Electric waves either of much higher, or much lower, frequency than light-waves would be capable of traversing the closed eyelids or the skull, or other material substance, without giving rise to sensible resonance, and therefore without absorption.

Now the existence of non-luminous radiation of low frequency, capable of traversing material objects of not too great thickness without sensible absorption, and of producing certain visual effects, and which, amongst numerous other sources, are known to be emitted by the brain and other parts of the body, is an observed fact. Rays subsequently found to possess these characteristics were discovered by R. Blondlot in 1893 in the radiation from a focus-tube employed for the production of Röntgen rays. They were observed in the stream of radiation from which the ordinary luminous rays had been absorbed by the passage of the stream through black paper. Blondlot called them N rays, and found them to be capable of traversing wood, metals, and other substances. I shall refer to them as Blondlot rays. He also observed them in the radiation from various sources of light,

[1] *Phil. Trans.*, 1897, A., p. 176.

such as a Nernst lamp, an Auer gas-burner, and even an ordinary ring gas-burner. Blondlot rays were found to increase the luminosity of sulphide of calcium and other phosphorescent screens which had previously been exposed to sunlight; and when made to impinge upon a small flame or spark, or upon a glowing wire or platinum plate, they were also found to increase the luminosity. This increase was observed on the side of the plate remote from the source of the rays as well as on the nearer side, although the plate was of sufficient thickness to absorb the radiation when cold. Delicate observations showed no increase in the temperature of the plate, due to the impact of the rays, so that they were capable of traversing it only when hot. The effects were very minute, and required a perceptible time for their development. The power of observing them was also found to vary greatly with different persons, some being quite incapable of detecting any effect at all. When the rays were concentrated by means of a quartz lens, their effects remained visible for some twenty minutes after the source of light from which they emanated had been extinguished, showing that the quartz had become a source of secondary radiation. Various metals were found to absorb and emit Blondlot rays in a similar manner. When the rays were made to impinge upon the eyeball in a dimly lighted room they were found to increase the visibility of the objects in the room, although the rays were not allowed to reach the latter. Their effect was therefore to increase the sensitiveness of the eye to ordinary light. This effect also was found to take time for its development, due to absorption of the rays by the eye and their subsequent emission, as was shown by experiments upon the eye of an ox which had been killed on the previous day. When wood, glass, caoutchouc, and other substances were compressed in a press they were found to emit Blondlot rays; and bodies in a state of internal strain, such as tempered steel and solidified crystalline sulphur, were found to provide a permanent source of these rays. For example, an eighteenth-century knife which had been in the possession of M. Blondlot's family for a period dating from shortly after its manufacture, and which had never been re-tempered, was found to emit the rays. Similar effects were produced temporarily by bending a strip of iron. Blondlot was unable to obtain any photographic effects from the rays, and his investigation of their physical properties led him to the conclusion that they consist of electric waves of greater wave-length, and therefore lower frequency, than visible light-waves.[1]

[1] R. Blondlot, *Comptes Rendus*, vol. cxxxvi., 1903, pp. 785, 1121, 1227, 1421; vol. cxxxvii., 1903, pp. 167, 686, 728, 831, 962.

The interest aroused by Blondlot's researches induced numerous observers to attempt their investigation; but many of them were unsuccessful in obtaining any results, which led some of them to regard the recorded effects as hallucinations on the part of the successful ones. Some of the most interesting of the successful observations were those made by A. Charpentier, the physiologist, who first noticed that the effect of Blondlot rays impinging upon a phosphorescent screen was apparently increased by the approach of the human body, and especially of a muscle or nerve, and that the effect was further increased by contraction of the muscle or stimulation of the nerve. The effects were found to persist after the interpolation of a thin aluminium screen, which showed that they were not due to rise of temperature, and Charpentier soon found that the observed effects were due to emission of these rays from the body. A number of successful experiments were made with rabbits and frogs, and it was found that the rays of physiological origin were in all respects identical with the Blondlot rays obtained from previously known sources. No emission of rays could be detected on the contraction of a tendon, but was observed on the compression of the supports and other bony portions. The presence of numerous nerves in the latter and their comparative scarcity in the tendons afford an explanation. When the compression was long continued, the emission of rays was found to diminish. Broca's centre, the nerve-centre in the brain which is regarded as the source of articulate speech, was found to possess exceptional emissive powers, and an increase was found to take place in the emission when the person under observation was speaking. Charpentier also believed that he was able to detect an increase in the emission from the cerebral nerve centres as a consequence of any mental exertion, such as concentration of the attention upon any object. He also found that the Blondlot rays gave rise to increased sensitiveness in the sense of smell when made to impinge either upon the nostril or upon the solutions of scent employed in the experiments. The latter effect was doubtless due to the absorption and subsequent emission of the radiation by the particles. The sensitiveness of the ear to sound was also found to be increased by Blondlot rays impinging upon, or close to, the ear, the impact being probably, as Charpentier observes, most effective upon the auditory nerve.[1]

It was shown by E. Meyer[2] that compression also gives rise to the emission of Blondlot rays in the case of plants, and that the

[1] A. Charpentier, *Comptes Rendus*, vol. cxxxvii., 1903, pp. 1049, 1277; vol. cxxxviii., 1904, pp. 584, 648.

[2] *Loc. cit.*, vol. cxxxviii., p. 101.

effect is diminished by the administration of chloroform and other anæsthetics. T. Becquerel[1] found that the emission from sulphide of calcium or quartz sand which had been previously exposed to sunlight was at first momentarily increased, and then gradually diminished, ultimately to extinction, by the application of chloroform and other anæsthetics. The emission reappeared when the anæsthetic vapour was dissipated by a current of air. Charpentier's results were confirmed by the investigations of G. Ballet,[2] who found that the emission of Blondlot rays is greatly decreased by certain pathological conditions of the nerves and muscles, as for example in cases of atrophy, and, generally, in cases of paralysis. Hysterical paralysis, however, was found to increase the emission. A. Broca[3] succeeded in locating certain portions of the cranium of men and dogs simply from observations on the emission of Blondlot rays.

Readers who desire to investigate the matter further will find references to many other researches on the Blondlot rays, indexed under the title "Rayons N," in the volumes of the *Comptes Rendus* for 1903 and 1904. The evidence here given, in my opinion, strongly supports the hypothesis which I have suggested to account for the phenomenon of thought-transference. This transference of thought by means of direct visual presentation is a very much simpler, and therefore in all probability a more primitive, method of communication than that of articulate speech; and it appears to me highly probable that it is a human faculty much more ancient than speech, and which has become gradually almost extinct owing to its supersession by the latter. The hypothesis of its retention, amongst some individuals at any rate, by uncivilised tribes, in the interior of Africa, for example, would account for the extraordinarily rapid transmission of intelligence across long distances which has appeared such a mystery to most African travellers. The same hypothesis affords a far more probable explanation of the communication which Lord Avebury has shown to be effected by ants, apparently through contact of the antennæ, than that which ascribes to the ant the intelligence requisite to devise a code of signals. It would, it seems to me, afford the most probable explanation of many of the mysteries of communication between animals, such, for example, as the simultaneous evolutions of large flocks of birds, and account satisfactorily for many of the otherwise inexplicable actions of the ordinary honey bee.[4] Moreover, if animals retain such powers, while we have almost entirely lost them, it appears probable that our thoughts, especially when representing concentrated will,

[1] *Loc. cit.*, p. 1159. [2] *Loc. cit.*, p. 524. [3] *Loc. cit.*, p. 1161.
[4] See, for example, Maeterlinck, *The Life of the Bee*.

would be capable of visual representation in their minds, and this would account very simply for many otherwise inexplicable instances of response by domestic animals to the apparently unexpressed will of the master. The mere mention of such instances as the sympathy between a horse and his rider or between a dog and his master will recall plenty of exemplifications to every lover of these animals.

INDEX

PRINTED BY NEILL AND CO., LTD., EDINBURGH.

·fm.aj12
4.50 nebx